Harvard University
Harvard University Biostatistics Working Paper Series

Year 2005 *Paper* 23

The Sensitivity and Specificity of Markers for Event Times

Tianxi Cai* Margaret S. Pepe† Thomas Lumley‡
Yingye Zheng** Nancy Swords Jenny††

*Harvard University, tcai@hsph.harvard.edu
†University of Washington, mspepe@u.washington.edu
‡University of Washington, tlumley@u.washington.edu
**Fred Hutchinson Cancer Research Center, yzheng@fhcrc.org
††University of Vermont, Nancy.Jenny@uvm.edu

http://biostats.bepress.com/harvardbiostat/paper23

The Sensitivity and Specificity of Markers for Event Times

Tianxi Cai

Department of Biostatistics, Harvard University, Boston, MA 02115

Margaret Sullivan Pepe, Thomas Lumley

Department of Biostatistics, University of Washington, Seattle, Washington 98195

Yingye Zheng

Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

and

Nancy Swords Jenny

Department of Pathology, College of Medicine , University of Vermont, Burlington, Vermont 05405

SUMMARY

The statistical literature on assessing the accuracy of risk factors or disease markers as diagnostic tests deals almost exclusively with settings where the test, Y, is measured concurrently with disease status D. In practice, however, disease status may vary over time and there is often a time lag between when the marker is measured and the occurrence of disease. One example concerns the Framingham Risk Score as a marker for the future risk of cardiovascular events, events that occur after the score is ascertained. To evaluate such a marker, one needs to take the time lag into account since the predictive accuracy may be higher when the marker is measured closer to the time of disease occurrence. We therefore consider inference for sensitivity and specificity functions that are defined as functions of time. Semi-parametric regression models are proposed. Data from a cohort study are used to estimate model parameters. One issue that arises in practice is that event times may be censored. In this research, we extend in several respects the work by Leisenring, Pepe and Longton (1997) that dealt only with parametric models for binary tests and uncensored data. We propose semi-parametric models that accommodate continuous tests and censoring. Asymptotic distribution theory for parameter estimates is developed and procedures for making statistical inference are evaluated with simulation studies. We illustrate our methods with data from the Cardiovascular Health Study, relating the Framingham risk score measured at enrollment to subsequent risk of cardiovascular events.

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. We define the time dependent TPR function:

$$\text{TPR}_{t,s}(y) = \text{pr}\left\{Y(s) \geq y \mid T - s = t\right\}, \tag{1·1}$$

where the time lag is t. For a subject that has an event at T, $\text{TPR}_{t,s}(y)$ is the probability of test positive with the marker at t time units prior to the event. Typically the TPR function will be a monotone decreasing function of t. The FPR, or 1 - specificity, relates to subjects without events, or at least event free by some suitably large time τ after the marker is measured. Therefore we define

$$\text{FPR}_{\tau,s}(y) = \text{pr}\left\{Y(s) \geq y \mid T - s > \tau\right\}, \tag{1·2}$$

These definitions are consistent with those used by Balasubramanian and Lagakos (2001), Leisenring, Pepe and Longton (1997) and Etzioni et al (1999) and are commonly used in practice. In breast cancer

screening research for example, $\tau = 2$ years is typically used. The rationale is that it can be assumed that if clinical disease does not emerge by 2 years after screening, the subject was free of subclinical disease at the time of screening. Heagerty, Pepe and Lumley (2000) define cumulative incidence based TPR and FPR functions:

$$\text{TPR}^{\text{CI}}_{t,s}(y) = \text{pr}\left\{Y(s) \geq y \mid T - s \leq t\right\}, \tag{1·3}$$

$$\text{FPR}^{\text{CI}}_{t,s,}(y) = \text{pr}\left\{Y(s) \geq y \mid T - s > t\right\}. \tag{1·4}$$

However, the definitions in (1·1) and (1·2) lead to more straightforward regression modeling procedures and are easier to interpret as functions of time than their cumulative incidence based counterparts. For extensive discussion see Pepe (2003, pages 259–65). Moreover, we can calculate (1·3) and (1·4) from (1·1) and (1·2) with knowledge of the event time distribution. Thus we concentrate on (1·1) and (1·2) here initially, and return to (1·3) and (1·4) in the example.

In this paper we consider models for the time dependent (TPR, FPR) functions in (1·1) and (1·2) 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. We develop our method in a simplified setting in the next section. The more general setting is discussed in section 3. Results from simulation studies described in section 4 suggest that the procedures work well in finite samples when the assumptions hold. In the second part of section 4 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 that it works better in females than in males. However, the score is not a very accurate predictor in any subgroup studied. We close in section 5 with a discussion of alternative approaches to the evaluation of markers for event time data.

2 Modelling and Estimation in A Simple Setting

We only consider marker measurements made prior to the event time, $Y(s)$ for $s \leq T$. The event time can be right censored by a censoring variable C that is independent of T conditional on the covariates. Define the observation time $X = \min(T, C)$ and $\delta = I(T \leq C)$. First, we consider the simple scenario with Y binary and measured at baseline, $s = 0$, for each person. The data for analysis are

$$\{(Y_i, \mathbf{Z}_i, X_i, \delta_i),\ i = 1, \ldots, n\}\ ,$$

where $\mathbf{Z}_i$ is the $p \times 1$ covariate vector. We assume the following models for the true and false positive rates:

$$\text{TPR}_{t,\mathbf{Z}_i} = \text{pr}(Y_i = 1 \mid T_i = t, \mathbf{Z}_i) = g_{\text{D}}\left\{\eta_{\boldsymbol{\alpha}_0}(t) + \boldsymbol{\beta}_0'\mathbf{Z}_i + h_0\right\},\ \text{ for } t \leq \tau, \tag{2·1}$$

$$\text{FPR}_{\tau,\mathbf{Z}_i} = \text{pr}(Y_i = 1 \mid T_i > \tau, \mathbf{Z}_i) = g_{\bar{\text{D}}}\left\{\mathbf{b}_0'\mathbf{Z}_i + c_0\right\}, \tag{2·2}$$

where g_{D} and $g_{\bar{\text{D}}}$ are specified inverse link functions. The dependence of the TPR on time is through the parametric function $\eta_{\boldsymbol{\alpha}_0}(t) = \boldsymbol{\alpha}_0'\boldsymbol{\eta}(t)$, where $\boldsymbol{\eta}$ is a vector of polynomials or spline basis functions. The false positive rate is not a function of time. The TPR and FPR rates (2·1) and (2·2) are mathematically distinct functions. Separate modeling of TPR and FPR is often undertaken in practice because the behavior of the marker may be very different in cases than in controls. See Carney et al (2003) for an example. However, it is also possible to specify (2·1) and (2·2) with shared parameters. Alternatively, one may assume that (2·1) holds for all t and drop (2·2) completely. Then the covariate specific FPR can be derived from (2·1) with $\text{FPR}_{\tau,\mathbf{Z}_i} = -\int_\tau^\infty \text{TPR}_{t,\mathbf{Z}_i} dS_{\mathbf{Z}_i}(t)/S_{\mathbf{Z}_i}(\tau)$, where $S_{\mathbf{Z}}(t)$ is the covariate specific survivor function for T. However, typically censoring does not allow estimation of TPR over $(0, \infty)$ so it is more appealing to assume (2·1) for $t < \tau$ and to supplement it with the FPR model (2·2). This is analogous to the approach taken by Hogan and Laird (1997) in the mixture model framework where they reserved a multinomial category for subjects who have not yet experienced an event by the end of the study.

Denote all the parameters in (2·1) and (2·2) by $\boldsymbol{\psi}_0$. Note that (2·1) models the distribution of Y_i conditional on $\{T_i, \mathbf{Z}_i\}$ when $T_i \leq \tau$ and (2·2) models the distribution of Y_i conditional on $\{T_i > \tau, \mathbf{Z}_i\}$. The likelihood of the data is therefore

$$\prod_{i=1}^{n} \{p_i(\boldsymbol{\psi}_0)\}^{Y_i} \{1 - p_i(\boldsymbol{\psi}_0)\}^{1-Y_i}\,, \tag{2·3}$$

where the conditional probability $p_i(\boldsymbol{\psi}_0)$ is

$$\begin{array}{ll} \mathrm{TPR}_{X_i,\mathbf{Z}_i} & \text{if } X_i \le \tau, \delta_i = 1, \\ \dfrac{-\int_{X_i}^{\tau} \mathrm{TPR}_{t,\mathbf{Z}_i} dS_{\mathbf{Z}_i}(t) + \mathrm{FPR}_{\tau,\mathbf{Z}_i} S_{\mathbf{Z}_i}(\tau)}{S_{\mathbf{Z}_i}(X_i)} & \text{if } X_i \le \tau, \delta_i = 0, \\ \mathrm{FPR}_{\tau,\mathbf{Z}_i} & \text{if } X_i > \tau, \end{array} \tag{2·4}$$

and $S_{\mathbf{Z}}(t) = \mathrm{pr}(T > t \mid \mathbf{Z})$. We can think of the first and third groups as cases and controls, respectively, while the case/control status of the second group, i.e. those censored in $[0, \tau]$, are unknown.

A consistent estimator of $\boldsymbol{\psi}_0$ can be obtained by using only data from subjects in the first and third groups. This is the approach taken by Leisenring et al. (1997). To incorporate observations from subjects who are censored before τ, one needs to estimate $S_{\mathbf{Z}}(\cdot)$ since the likelihood contributions from these subjects involve $S_{\mathbf{Z}}(\cdot)$. To this end, we assume a proportional hazards model for $T_i \mid \mathbf{Z}_i$:

$$\lambda(t \mid \mathbf{Z}_i) = \lambda_0(t) \exp(\boldsymbol{\gamma}_0' \mathbf{Z}_i), \tag{2·5}$$

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. Then $S_{\mathbf{Z}_i}(t) = \mathrm{pr}(T \ge t \mid \mathbf{Z}_i) = \exp\{-\Lambda_0(t) \exp(\boldsymbol{\gamma}_0' \mathbf{Z}_i)\}$, where $\Lambda_0(\cdot)$ is the baseline cumulative hazard function. Under the Cox model (2·5), the survival function $S_{\mathbf{Z}}(t)$ can be consistently estimated as

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

where $\hat{\boldsymbol{\gamma}}$ is the maximum partial likelihood estimator based on $\{X_i, \delta_i, \mathbf{Z}_i;\ i = 1, \ldots, n\}$, and $\hat{\Lambda}_0(t)$ is the Breslow estimate of the cumulative baseline hazard function $\Lambda_0(t)$ (Fleming and Harrington, 1991). Plugging in the estimated survival function to (2·4), we obtain approximate conditional probabilities, $\widehat{p}_i(\boldsymbol{\psi})$. Then $\boldsymbol{\psi}_0$ can be estimated by maximizing the approximated likelihood function, or equivalently, as the solution to the approximated score equation,

$$\sum_{i=1}^{n} \frac{\frac{\partial \widehat{p}_i(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}}}{\widehat{p}_i(\boldsymbol{\psi})\{1 - \widehat{p}_i(\boldsymbol{\psi})\}} \{Y_i - \widehat{p}_i(\boldsymbol{\psi})\} = 0. \tag{2·6}$$

We note that other regression models for T can be used. Estimation of $\boldsymbol{\psi}_0$ only requires a consistent estimate of $S_{\mathbf{z}}(t)$ and does not rely on the proportional hazards model assumption.

3 General Semi-parametric Framework

3·1 *Model Assumptions*

We now generalize to a continuous marker, Y, and allow measurements at various times s_{ik}, $k = 1, ..., K_i$, for subject i. To deal with multiple marker measurements, each observation is taken as a unit and its time origin is reset to s_{ik}. We model the marginal probability associated with $(Y_{ik}, T_{ik}, \mathbf{Z}_i)$, where Y_{ik} is the marker measured at s_{ik} and $T_{ik} = T_i - s_{ik}$ is the time lag between the measurement time and the occurrence of the event. Consider the following marginal probability models:

$$\text{TPR}_{t,s_{ik},\mathbf{Z}_i}(y) = \text{pr}(Y_{ik} \geq y \mid T_{ik} = t, \mathbf{Z}_i, s_{ik}) = g_{\text{D}}\left\{\eta_{\boldsymbol{\alpha}_0}(t, s_{ik}) + \boldsymbol{\beta}_0'\mathbf{Z}_i + h_0(y)\right\},\ t \leq \tau, \tag{3·1}$$

$$\text{FPR}_{\tau,s_{ik},\mathbf{Z}_i}(y) = \text{pr}(Y_{ik} \geq y \mid T_{ik} > \tau, \mathbf{Z}_i, s_{ik}) = g_{\bar{\text{D}}}\left\{\xi_{\boldsymbol{a}_0}(s_{ik}) + \mathbf{b}_0'\mathbf{Z}_i + c_0(y)\right\}, \tag{3·2}$$

where h_0 and c_0 are baseline functions of the threshold y that are completely unspecified. The TPR function is allowed to depend on both the time lag T_{ik} and the measurement time s_{ik}. The FPR function may only depend on the measurement time. 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. In many applications the sensitivity $\text{TPR}_{t,s,\mathbf{z}}(\cdot)$ of the marker depends only on the time lag t. However, in some applications the absolute time of the marker measurement 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 previous section where s_{ik} is a constant for all observations (e.g. $s_{ik} = 0$), we set $\xi_{\boldsymbol{a}_0}(s) = 0$.

The models written in (2·1) and (2·2) assume that covariate effects are additive and do not depend on the thresholding value y. However, one can allow covariate effects to vary with the thresholding value by including interactions between covariates and parametric functions of y. This is similar to relaxing the proportional hazards assumption of the Cox model for failure time data by including interactions between covariate and parametric functions of time. Our estimating procedures apply to the more general model, but to keep notation simple we work with the simpler model here.

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 and $\eta_{\boldsymbol{\alpha}_0}(\cdot, \cdot)$ and $\xi_{\boldsymbol{a}_0}(\cdot)$ quantify time effects. Note that, 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 (3·1) and (3·2) can be represented as:

$$h_0(Y_{ik}) = -\boldsymbol{\eta}_{\boldsymbol{\alpha}_0}(T_{ik}, s_{ik}) - \boldsymbol{\beta}_0'\mathbf{Z}_i + \varepsilon_{\mathrm{D}ik}, \qquad \text{if } T_{ik} \le \tau,$$
$$c_0(Y_{ik}) = -\boldsymbol{\xi}_{\boldsymbol{a}_0}(s_{ik}) - \mathbf{b}_0'\mathbf{Z}_i + \varepsilon_{\bar{\mathrm{D}}ik}, \qquad \text{if } T_{ik} > \tau,$$

where $\mathrm{pr}(\varepsilon_{\mathrm{D}ik} \ge y) = g_{\mathrm{D}}(y)$ and $\mathrm{pr}(\varepsilon_{\bar{\mathrm{D}}ik} \ge y) = g_{\bar{\mathrm{D}}}(y)$.

3·2 *Estimating the Model Components*

Let $\boldsymbol{\psi}_0 = \left\{\mathbf{H}_0(\cdot) = [h_0(\cdot), c_0(\cdot)]', \boldsymbol{\theta}_0 = [\boldsymbol{\alpha}_0', \boldsymbol{\beta}_0', \boldsymbol{a}_0', \mathbf{b}_0']'\right\}$ denote all the unknown parameters. We base inference for $\boldsymbol{\psi}_0$ on indicator variables $I(Y_{ik} \ge y)$ since models (3·1) and (3·2) essentially relate its conditional expectation $p_{ik}(y; \boldsymbol{\psi}_0)$ to the parameters of interest $\boldsymbol{\psi}_0$. Similar to the binary case, the probability $p_{ik}(y; \boldsymbol{\psi})$ depends on the case/control/censored status of the observation and can be derived from the models (3·1) and (3·2) for subjects who have an event before τ, who are censored before τ and who do not have an event before τ, respectively. To estimate the non-parametric baseline functions, $\mathbf{H}_0(y)$, we consider the marginal binomial likelihood based on $I(Y_{ik} \ge y)$ and solve the corresponding score equation

$$\sum_{i=1}^{n}\sum_{k=1}^{K_i} \mathbf{W}_{ik}^{\mathbf{H}}(y; \boldsymbol{\psi}) \left\{I(Y_{ik} \ge y) - p_{ik}(y; \boldsymbol{\psi})\right\} = 0\,, \qquad \text{for each } \; y \in [\mathfrak{l}, \mathfrak{u}]\,, \tag{3·3}$$

where $\mathbf{W}_{ik}^{\mathbf{H}}(y; \boldsymbol{\psi}) = \frac{\rho^{1-\delta_{ik}} \frac{\partial}{\partial \mathbf{H}} p_{ik}(y; \boldsymbol{\psi})}{p_{ik}(y; \boldsymbol{\psi})\{1 - p_{ik}(y; \boldsymbol{\psi})\}}$, ρ is a pre-specified non-negative weight, $\delta_{ik} = \delta_i I(X_i - s_{ik} \le \tau) + I(X_i - s_{ik} > \tau)$ 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

$$\sum_{i=1}^{n}\sum_{k=1}^{K_i} \int_{\mathfrak{l}}^{\mathfrak{u}} \mathbf{W}_{ik}^{\boldsymbol{\theta}}(y; \boldsymbol{\psi}) \Big\{I(Y_{ik} \ge y) - p_{ik}(y; \boldsymbol{\psi})\Big\} d\hat{v}(y) = 0\,, \tag{3·4}$$

where $\mathbf{W}_{ik}^{\boldsymbol{\theta}}(y; \boldsymbol{\psi}) = \frac{\rho^{1-\delta_{ik}} \frac{\partial}{\partial \boldsymbol{\theta}} p_{ik}(y; \boldsymbol{\psi})}{p_{ik}(y; \boldsymbol{\psi})\{1 - p_{ik}(y; \boldsymbol{\psi})\}}$, $\hat{v}(\cdot)$ 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 (3·3) and (3·4) simultaneously to estimate the parameters in the models (3·1) and (3·2).

Observe that we now include in the estimating equations a weighting factor ρ that dictates the extent to which the censored observations in $[0, \tau]$ enters into the analysis (we have $\rho^{1-\delta_{ik}} = 1$ for known cases and controls because $1 - \delta_{ik} = 0$ for them). If ρ is set to 0, then censored observations

are excluded entirely from the analysis which corresponds to the Leisenring et al (1997) approach. The score equation given in (2·6) for the binary case essentially sets ρ to 1. Increasing ρ allows censored observations to have more influence on estimation. A variety of values for ρ are investigated later in simulation studies. Again, 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. In summary, we propose a two step approach to estimating $\boldsymbol{\psi}_0$:

(1) 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 (3·3) and (3·4) when $\rho = 0$, i.e. when the censored observations with $X_i - s_{ik} \leq \tau$ and $\delta_i = 0$ are ignored. $\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, h_0(\cdot)\}$ are estimated using only the cases, observations with $X_i - s_{ik} \leq \tau$ and $\delta_i = 1$, and $\{\boldsymbol{a}_0, \mathbf{b}_0, c_0(\cdot)\}$ are estimated using only the controls, i.e. observations with $X_i - s_{ik} > \tau$. We show in Appendix A that $\widetilde{\boldsymbol{\psi}}$ is a consistent estimator of $\boldsymbol{\psi}_0$ for $y \in [\mathfrak{l}, \mathfrak{u}]$.

(2) Estimation of $\boldsymbol{\psi}_0$ when $\rho > 0$.

To include observations with $\delta_{ik} = 0$ in estimating $\boldsymbol{\psi}_0$, we set $\rho > 0$ and approximate $p_{ik}(y;\boldsymbol{\psi})$ for those censored subjects by using the estimated survivor function $S_{\mathbf{z}}(\cdot)$ as in the binary case assuming a proportional hazards model (2·5). Then we solve (3·3) and (3·4) using the weight functions $\mathbf{W}^{\mathbf{H}}_{ik}(y;\boldsymbol{\psi})$ and $\mathbf{W}^{\boldsymbol{\theta}}_{ik}(y;\boldsymbol{\psi})$ evaluated at $\widetilde{\boldsymbol{\psi}}$. Let $\widehat{\boldsymbol{\psi}} = \{\widehat{h}(\cdot), \widehat{c}(\cdot), \widehat{\boldsymbol{\theta}}\}$ denote the final estimator for $\boldsymbol{\psi}_0$.

3·3 *Inference in Large Samples*

We show in appendix A that $\widehat{\boldsymbol{\psi}}$ is unique for large n and is consistent. Furthermore, $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically equivalent to $n^{-\frac{3}{2}}\mathbb{A}^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{U}_{ij}$, where $\mathbb{A}$ and $\mathbf{U}_{ij}$ are defined in appendix A and B, respectively. It follows from properties of U-statistics (Serfling, 1980) that $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is approximately normal with mean $\mathbf{0}$ and variance $\boldsymbol{\Sigma} = n^{-3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1,k\neq j}^{n}(\mathbf{U}_{ij}+\mathbf{U}_{ji})(\mathbf{U}_{ik}+\mathbf{U}_{ki})'$. Now, let $\widehat{\mathbb{A}}$ and $\widehat{\boldsymbol{\Sigma}}$ be the matrices obtained by replacing all the theoretical quantities in $\mathbb{A}$ and $\boldsymbol{\Sigma}$ with their empirical counterparts. Then, the covariance matrix of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ can be approximated by $\widehat{\mathbb{A}}^{-1}\widehat{\boldsymbol{\Sigma}}\widehat{\mathbb{A}}^{-1}$.

We now turn to inference about the TPR and FPR functions, which depend on time and on covariates $\mathbf{z}$. Substituting $\{\widehat{\boldsymbol{\theta}}, \widehat{h}(y), \widehat{c}(y)\}$ into (3·1) and (3·2), we propose estimating $\text{TPR}_{t,s,\mathbf{z}}(y)$ and $\text{FPR}_{\tau,s,\mathbf{z}}(y)$

as:

$$\widehat{\text{TPR}}_{t,s,\mathbf{z}}(y) = g_{\text{D}}\left\{\eta_{\widehat{\boldsymbol{\alpha}}}(t,s) + \widehat{\boldsymbol{\beta}}'\mathbf{z} + \widehat{h}(y)\right\}, \quad \widehat{\text{FPR}}_{\tau,s,\mathbf{z}}(y) = g_{\bar{\text{D}}}\left\{\xi_{\widehat{\boldsymbol{\alpha}}}(s) + \widehat{\mathbf{b}}'\mathbf{z} + \widehat{c}(y)\right\}.$$

To obtain pointwise and simultaneous confidence intervals for the TPR and FPR functions, we show in appendix B that the process $\mathbf{Q}(y;t,s,\mathbf{z}) \equiv n^{\frac{1}{2}}[g_{\text{D}}^{-1}\{\widehat{\text{TPR}}_{t,s,\mathbf{z}}(y)\} - g_{\text{D}}^{-1}\{\text{TPR}_{t,s,\mathbf{z}}(y)\}, g_{\bar{\text{D}}}^{-1}\{\widehat{\text{FPR}}_{\tau,s,\mathbf{z}}(y)\} - g_{\bar{\text{D}}}^{-1}\{\text{FPR}_{\tau,s,\mathbf{z}}(y)\}]'$ is asymptotically equivalent to

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

where $\mathcal{H}_{ij}(y)$ is defined in (B.1). This 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. This technique avoids the need to derive explicit analytic expressions for variance-covariance processes, which seem intractable in our setting. Moreover, relative to other re-sampling methods such as the bootstrap, the computational burden is minimal. A detailed description of the procedure for constructing confidence bands based on the re-sampling method can be found in Cai & Pepe (2002).

4 Simulation Studies and Example

4·1 *Simulation Studies*

Simulation studies were performed to examine the finite sample properties of the estimation procedures proposed in the previous sections and to investigate the impact of ρ on their efficiencies. The results suggest that our methods provides reasonably unbiased estimates of both the model parameters and (TPR, FPR) functions. The choice of ρ affects the efficiency of all parameters of interest, but seems to have most impact on $\widehat{\mathbf{b}}$, less on $\widehat{\beta}$ and very little on $\widehat{\alpha}$ in the settings we considered. Recall that when $\rho = 0$, censored observations are ignored for estimation. By including censored observations, i.e. using $\rho > 0$, we find that the estimates are almost always more precise than those calculated with $\rho = 0$. See Cai et al (2003) for details on simulation results.

To choose an optimal weighting of censored data in practice, 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 squared 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 $\sum_{j=1}^{P} \sqrt{\text{var}(\widehat{\theta}_j)/\widehat{\theta}_j^2}$, where P is the dimension of $\boldsymbol{\theta}$ and $\widehat{\theta}_j$ is the jth component of $\widehat{\boldsymbol{\theta}}$.

4·2 *Example: The CHS Study*

The Cardiovascular Heath Study (CHS) is a population based observational prospective study of elderly adults (age $\geq$ 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) between 65 and 95 years old who were free of clinical cardiovascular events at enrollment. There were 585 (14.7%) who experienced a cardiovascular event during the study. Follow up on subjects without cardiovascular events averaged 6.75 years (sd = 1.58 years). Here we consider the Framingham Risk Score (FR-score) as a predictor for cardiovascular events. The 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.

The FR-score was evaluated for all subjects at enrollment, and we only include this baseline assessment in this analysis. Thus $s = 0$ for all observations. We considered gender and medication for hypertension as covariates that might have a substantial influence on the predictive accuracy of the FR-score. We fit the following models for the time dependent TPR and FPR functions:

$$\text{TPR}_{t,\mathbf{Z}}(y) = 1 - \Phi\{\alpha_1 t + \alpha_2 t^2 + \beta_1 Z_1 + \beta_2 Z_2 + h_0(y)\}$$
$$\text{FPR}_{\tau,\mathbf{Z}}(y) = 1 - \Phi\{b_1 Z_1 + b_2 Z_2 + c_0(y)\}$$

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 (3·1) and (3·2) 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 which would imply that they were subsequently at risk for cardiovascular events, we censor them at the end of the

trial. Since $1 - S_{\mathbf{Z}}(t)$ represents the marginal probability of a cardiovascular events by time t, it is appropriate to include them as definite non-events in the estimation of $S_{\mathbf{Z}}(t)$ (Pepe and Mori, 1993).

The estimated regression coefficients and their estimated standard errors are shown in Table 1. The estimated FPR and TPR functions for different groups at $t = 1$ year and $t = 5$ years are shown in Figure 1 and their corresponding ROC curves are presented in Figure 2. The lack of a gender effect on the FPR functions ($b_2 = 0$) indicate that in order to control the FPR at a particular value, the same threshold value y can be used for males and females. The positive value of β_2 indicates however that for a given threshold value, the FR-score appears to have higher sensitivity in females than in males for detecting subsequent cardiovascular events but similar specificity. The estimated FPR and TPR functions shown in Figure 1 illustrate this. Medication use appears to be associated with higher false positive and true positive rates since both coefficient estimates b_1 and β_1 are negative. The medication effect on the TPR is less than that on the FPR. This translates into the ROC curve for those on medication being less than for those who are not on medication, as illustrated in Figure 2. In other words, the FR-score is better at distinguishing between subjects who will and will not have a cardiovascular event when they are not on medication. We have noted that given the same false positive rate, the corresponding true positive rate is substantially higher in females than in males. As shown in Table 2, the area under the ROC curve (AUC) in females is about 0.07 (sd = 0.02) higher than that in males in both medication groups at year 1. The gender effect is statistically significant. However, the medication effect is not. The AUC in subjects on medication is only about 0.03 (sd = 0.02) lower than that in subjects not on medication (for both male and female) at year 1.

Perhaps the most important and disappointing result of our analysis however is that the FR-score is a very inaccurate marker for cardiovascular events. The ROC curves in Figure 2 demonstrate that the benefit of a high TPR can only be achieved at the expense of an accompanying high FPR and vice versa. It does not have adequate sensitivity and specificity (at any threshold) for accurate individual level prediction of cardiovascular events. Figure 3 displays the estimated TPR functions for events at at $t = 1$ year and 5 years after the FR-score is measured in male subjects who are not 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 men, the threshold criterion FR-score > 10 which identifies 45% of subjects with events at 1 year after enrollment, identifies only 36% of subjects with events at 5 years. The FPR functions for those

still without events at 7 years after enrollment is also shown in Figure 3. Of those subjects, 30% also meet criterion that FPR-score > 10.

We next calculate the cumulative incidence based TPR and FPR functions defined in section 1 and derive the area under the corresponding ROC curves. Estimators for $\text{TPR}^{\text{CI}}_{t,s,\mathbf{z}}(\cdot)$ and $\text{FPR}^{\text{CI}}_{t,s,\mathbf{z}}(\cdot)$ can be derived from our proposed models in (3·1) and (3·2) noting the following identities:

$$\text{TPR}^{\text{CI}}_{t,s,\mathbf{z}}(y) = \frac{-\int_0^t g_{\text{D}}\left\{\eta\boldsymbol{\alpha}_0(u,s) + \boldsymbol{\beta}_0'\mathbf{z} + h_0(y)\right\} dS_{\mathbf{z}}(u+s)}{1 - S_{\mathbf{z}}(t+s)},$$

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

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. The results are presented in Figure 4 and Table 3. At any given positivity threshold, both the true and false positive rates of the FR-score are higher for $t = 1$ year since enrollment than for $t = 5$ years since enrollment. The FR-score is better at distinguishing subjects who fail within 1 year from those do not than distinguishing subjects who fail within 5 years from those do not. In particular, for female subjects who are on hypertension medication, the AUC for the cumulative incidence ROC curve is 0.68 (sd= 0.03) when $t = 1$ year and is 0.61 (sd=0.02) when $t = 5$ years. Comparing the AUC between 1 year and 5 years for this subgroup, we have a difference of 0.07 with standard error 0.02. Shown also in Figure 4 are the non-parametric estimates (Heagerty et al, 2000) of the ROC curves using data from each subgroup. The nonparametric estimates are reasonably close to the model based estimates suggesting that the assumed semi-parametric models are reasonable.

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

There are several directions in which the methodology proposed here should be enhanced. First, procedures to assess model fit and to assist in model selection need development. Although the models are semiparametric, there is an assumption that covariate effects are additive on the scale of the link function and we did not test the adequacy of this assumption in our data analysis. Second, we have restricted to time independent covariates. It would be of interest to generalize the models to include time varying covariates. This is straightforward if covariates are external but more complicated when covariates are internal in the sense defined by Kalbfleisch and Prentice (2002). Third, we have restricted to biomarkers measured prior to the occurrence of the event. However in some applications, such as in infectious disease research, there would be interest in using the biomarker to detect the event as soon as possible *after* T. Fourth, we do not allow censoring of T to depend on the marker Y conditional on the covariates. To do so would induce bias into the estimated true and false positive rates, commonly referred to as 'verification bias' in the diagnostic testing literature (Begg and Greenes, 1983). Studies that seek to evaluate the sensitivity and specificity of a marker should be designed so that follow up does not depend on the marker. However, extensions that relax this requirement somewhat could be possibly developed using inverse probability weighting for example (Robins et al, 1995). Finally, our methods can be used to assess the relative performance of two markers. One would fit separate models for each of the markers and plot the induced ROC curves. Comparisons could then be made using AUC statistics for example. It would be interesting to explore this further with real data.

ACKNOWLEDGEMENT

The authors are grateful to reviewers for constructive comments that lead to substantial improvement to the paper. Support for this research was provided by NIH grants GM-54438 and AI29168. Participating Institutions and Principal Investigators for CHS are: Wake Forest University School of

Medicine: Gregory L. Burke MD. Wake Forest University-ECG Reading Center: Pentti M. Rautaharju MD PhD. University of California, Davis: John Robbins MD MHS. The Johns Hopkins University: Linda P. Fried MD MPH. The Johns Hopkins University-MRI Reading Center: Nick Bryan MD PhD, Norman J. Beauchamp MD. University of Pittsburgh: Lewis H. Kuller, MD DrPH. University of California, Irvine-Echocardiography Reading Center (baseline): Julius M. Gardin MD. Georgetown Medical Center-Echocardiography Reading Center (follow-up): John S. Gottdiener MD. New England Medical Center, Boston-Ultrasound Reading Center: Daniel H. O'Leary MD. University of Vermont-Central Blood Analysis Laboratory: Russell P. Tracy PhD. University of Arizona, Tucson-Pulmonary Reading Center: Paul Enright MD. Retinal Reading Center-University of Wisconson: Ronald Klein MD. University of Washington-Coordinating Center: Richard A. Kronmal PhD. NHLBI Project Office: Jean Olson MD MPH. The CHS research was supported by contracts N01-HC-85079 through N01-HC-85086, N01-HC-35129, and N01-HC-15103 from the National Heart, Lung, and Blood Institute.

APPENDIX

A UNIQUENESS AND CONSISTENCY

We first show that $\widetilde{\boldsymbol{\psi}}$ is a consistent estimate of $\boldsymbol{\psi}_0$ for $y \in [\mathfrak{l}, \mathfrak{u}]$. To this end, let $[\widetilde{h}(y;\boldsymbol{\theta}), \widetilde{c}(y;\boldsymbol{\theta})]'$ denote the solution to (3·3) at a given $\boldsymbol{\theta}$ and

$$\ell(y;\boldsymbol{\psi}) = \sum_{\delta_{ik}=1} \sum_{k=1}^{K_i} \Big[I(Y_{ik} \geq y) \log p_{ik}(y;\boldsymbol{\psi}) + \{1 - I(Y_{ik} \geq y)\} \log \{1 - p_{ik}(y;\boldsymbol{\psi})\} \Big].$$

It is easy to see that $\widetilde{h}(y;\boldsymbol{\theta}) = \text{argmax}_{h(y)}\, \ell(y;\boldsymbol{\psi})$, $\widetilde{c}(y;\boldsymbol{\theta}) = \text{argmax}_{c(y)}\, \ell(y;\boldsymbol{\psi})$ and $\widetilde{\boldsymbol{\theta}} = \text{argmax}_{\boldsymbol{\theta}}\, \widetilde{Q}(\boldsymbol{\theta})$, 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 $\mathfrak{D}_\theta$. 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 $\mathcal{F} = \{I(X \leq \tau), \delta, \min(X, \tau), s, \mathbf{Z}\}$, $\widetilde{h}(y) = \widetilde{h}(y;\widetilde{\boldsymbol{\theta}})$ and $\widetilde{c}(y) = \widetilde{c}(y;\widetilde{\boldsymbol{\theta}})$. 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 $\mathcal{K}_0 = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} K_i \delta_{ik}$. 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 $\mathcal{F}_{11} \in \mathfrak{D}_{\mathcal{F}}$ where $\mathfrak{D}_{\mathcal{F}}$ is the domain of $\mathcal{F}_{11}$. It follows from the monotonicity of $g_{\text{D}}(\cdot)$ and $g_{\bar{\text{D}}}(\cdot)$ that

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

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

for $y \in \Re_v$ and any $(x, s, \mathbf{z})$ in the subspace of $\mathfrak{D}_{\mathcal{F}}$ restricting to $\Delta = 0$. Without loss of generality, we assume that $\mathfrak{D}_{\mathcal{F}}$ is not degenerate. It follows that (A.1) and (A.2) holds if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, $h(y) = h_0(y)$, 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 $\boldsymbol{\theta} \in \mathfrak{D}_\theta$. 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 $\widehat{p}$, $\widehat{\mathbf{W}}^{\mathbf{H}}_{ik}$, $\widehat{\mathbf{W}}^{\boldsymbol{\theta}}_{ik}$ be p, $\mathbf{W}^{\mathbf{H}}_{ik}$, $\mathbf{W}^{\boldsymbol{\theta}}_{ik}$, with $S_{\mathbf{Z}_i}(\cdot)$ replaced with $\widehat{S}_{\mathbf{Z}_i}(\cdot)$, respectively. Let $\widehat{\mathbf{H}}(y;\boldsymbol{\theta}) = [\widehat{h}(y;\boldsymbol{\theta}), \widehat{c}(y;\boldsymbol{\theta})]'$ denote the solution to

$$\sum_{i=1}^{n}\sum_{k=1}^{K_i} \widehat{\mathbf{W}}^{\mathbf{H}}_{ik}(y;\widetilde{\boldsymbol{\psi}}) \{I(Y_{ik} \geq y) - \widehat{p}_{ik}(y;\boldsymbol{\psi})\} = 0\,, \qquad \text{for any } \; y \in [\mathfrak{l}, \mathfrak{u}]\,. \tag{A.3}$$

given $\boldsymbol{\theta}$ and let $\widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}} = \left\{\widehat{\mathbf{H}}(\cdot;\boldsymbol{\theta});\boldsymbol{\theta}\right\}$. Then $\widehat{\boldsymbol{\theta}}$ is the solution to

$$\sum_{i=1}^{n}\sum_{k=1}^{K_i} \int_{\mathfrak{l}}^{\mathfrak{u}} \widehat{\mathbf{W}}^{\boldsymbol{\theta}}_{ik}(y;\widetilde{\boldsymbol{\psi}}) \left\{I(Y_{ik} \geq y) - \widehat{p}_{ik}(y;\widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}})\right\} d\hat{v}(y) = 0\,. \tag{A.4}$$

Let $\mathbf{V}^{\mathbf{H}}(y;\boldsymbol{\psi})$ and $\mathbf{V}(\boldsymbol{\theta})$ denote the left hand side of (A.3) and (A.4), respectively. It follows from the strong law of large numbers and the uniform consistency of $\widehat{S}_{\mathbf{z}}(\cdot)$ that for any $\mu \geq 0$, $y \in [\mathfrak{l}, \mathfrak{u}]$, $\boldsymbol{\theta} \in \mathfrak{D}_\mu = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \mu\}$, and sufficiently large ϵ,

$$\mathbf{V}^{\mathbf{H}}(y;\boldsymbol{\psi}_{+\epsilon}) > 0\,, \quad \mathbf{V}^{\mathbf{H}}(y;\boldsymbol{\psi}_{-\epsilon}) < 0 \tag{A.5}$$

where $\boldsymbol{\psi}_{\pm\epsilon} = \{\mathbf{H}(\cdot) \pm \epsilon;\boldsymbol{\theta}\}$. This, coupled with the monotonicity and continuity of g and $g_{\bar{\mathfrak{d}}}$, implies that there exists a unique $\widehat{\mathbf{H}}(\cdot;\boldsymbol{\theta})$ such that $\mathbf{V}^{\mathbf{H}}\left\{y;\widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}}\right\} = 0$, where $\widehat{\boldsymbol{\psi}}_{\boldsymbol{\theta}} = \left\{\widehat{\mathbf{H}}(\cdot;\boldsymbol{\theta});\boldsymbol{\theta}\right\}$. Since (A.5) 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 by the consistency of $\widetilde{\boldsymbol{\psi}}$, $\widehat{\mathbf{H}}(\cdot;\boldsymbol{\theta}_0)$ and a uniform law of large numbers (Pollard, 1990), $n^{-1}\partial\mathbf{V}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}' \to \mathbb{A}$ almost surely, where $\mathbb{A} = -\int_{\mathfrak{l}}^{\mathfrak{u}}[\mathbb{W}^{\boldsymbol{\theta}\boldsymbol{\theta}}(y) - \mathbb{W}^{\mathbf{H}\boldsymbol{\theta}}(y)'\{\mathbb{W}^{\mathbf{H}\mathbf{H}}(y)\}^{-1}\mathbb{W}^{\mathbf{H}\boldsymbol{\theta}}(y)]dv(y)$, $\mathbb{W}^{\boldsymbol{\theta}\boldsymbol{\theta}}(y) = E\{\mathbb{W}^{[02]}_{ik}(y;\boldsymbol{\psi}_0)\}$, $\mathbb{W}^{\mathbf{H}\boldsymbol{\theta}}(y) = E\{\mathbb{W}^{[11]}_{ik}(y;\boldsymbol{\psi}_0)\}$, $\mathbb{W}^{\mathbf{H}\mathbf{H}}(y) = E\{\mathbb{W}^{[20]}_{ik}(y;\boldsymbol{\psi}_0)\}$, $\mathbb{W}^{[ml]}_{ik}(y;\boldsymbol{\psi}) = \frac{\frac{\partial p_{ik}}{\partial \mathbf{H}}(y;\boldsymbol{\psi})^{\otimes m}\{\frac{\partial p_{ik}}{\partial \boldsymbol{\theta}}(y;\boldsymbol{\psi})^{\otimes l}\}'}{p_{ik}(y;\boldsymbol{\psi})\{1-p_{ik}(y;\boldsymbol{\psi})\}}$, 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 $n^{-\frac{1}{2}}\mathbf{V}(\boldsymbol{\theta}_0)$ is asymptotically equivalent to

$$n^{-\frac{1}{2}}\sum_{\delta_{ik}=0}\int_{\mathfrak{l}}^{\mathfrak{u}}\left[\int_{X_{ik}}^{\tau} g_{\mathrm{D}}\left\{\boldsymbol{\eta}_{\boldsymbol{\alpha}_0}(u,s_{ik})+\boldsymbol{\beta}_0'\mathbf{Z}_i+h_0(y)\right\}d\left\{\frac{S_{\mathbf{Z}_i}(u+s_{ik})}{S_{\mathbf{Z}_i}(X_i)}-\frac{\widehat{S}_{\mathbf{Z}_i}(u+s_{ik})}{\widehat{S}_{\mathbf{Z}_i}(X_i)}\right\}\right.$$
$$\left.+\,g_{\bar{\mathrm{D}}}\left\{\xi_{\boldsymbol{a}_0}(s_{ik})+\mathbf{b}_0'\mathbf{Z}_i+c_0(y)\right\}\left\{\frac{S_{\mathbf{Z}_i}(\tau+s_{ik})}{S_{\mathbf{Z}_i}(X_i)}-\frac{\widehat{S}_{\mathbf{Z}_i}(\tau+s_{ik})}{\widehat{S}_{\mathbf{Z}_i}(X_i)}\right\}\right]dv(y)+n^{-\frac{1}{2}}\sum_{i,k}\int_{\mathfrak{l}}^{\mathfrak{u}}\overline{\mathbf{W}}_{ik}^{\boldsymbol{\theta}}(y)e_{ik}(y)dv(y)$$

where $X_{ik}=X_i-s_{ik}$, $\overline{\mathbf{W}}_{ik}^{\boldsymbol{\theta}}(y)=\mathbf{W}_{ik}^{\boldsymbol{\theta}}(y;\boldsymbol{\psi}_0)-\mathbb{W}^{\mathbf{H}\boldsymbol{\theta}}(y)'\{\mathbb{W}^{\mathbf{H}\mathbf{H}}(y)\}^{-1}\mathbf{W}_{ik}^{\mathbf{H}}(y;\boldsymbol{\psi}_0)$. Recall that $\widehat{S}_{\mathbf{z}}(t)=\exp\{-\hat{\Lambda}_0(t)\exp(\hat{\boldsymbol{\gamma}}'\mathbf{z})\}$, 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 $x\le u\le\tau$:

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

where $\mathfrak{s}_i(u,x;\mathbf{z})=\frac{S_{\mathbf{z}}(u)}{S_{\mathbf{z}}(x)}\exp(\boldsymbol{\gamma}_0'\mathbf{z})[\ell_i'\{\mathbf{O}(u;\mathbf{z})-\mathbf{O}(x;\mathbf{z})\}+\int_x^u\frac{dM_i(s)}{\mathbb{S}^{(0)}(s)}]$, $\ell_i=\int_0^\infty\{\mathbf{Z}_i-\frac{\mathbb{S}^{(1)}(t)}{\mathbb{S}^{(0)}(t)}\}dM_i(t)$, $M_i(t)=I(X_i\le t)\delta_i-\int_0^t I(X_i\ge s)e^{\boldsymbol{\gamma}_0'\mathbf{Z}_i}d\Lambda_0(s)$, $\mathbb{S}^{(k)}(t)=E\{I(X\ge t)\exp(\boldsymbol{\gamma}_0'\mathbf{Z})\mathbf{Z}^{\otimes k}$ and $\mathbf{O}(u;\mathbf{z})=\left[E\left\{\delta\frac{\mathbb{S}^{(2)}(X)\mathbb{S}^{(0)}(X)-\mathbb{S}^{(1)}(X)^{\otimes 2}}{\mathbb{S}^{(0)}(X)^2}\right\}\right]^{-1}\left[\Lambda_0(u)\mathbf{z}+E\left\{\frac{\delta I(X\le u)\mathbb{S}^{(1)}(X)}{\mathbb{S}^{(0)}(X)^2}\right\}\right]$. As a consequence,

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

where $\mathbf{U}_{ij}=\int_{\mathfrak{l}}^{\mathfrak{u}}\{\mathcal{E}_{ij}(y;\overline{\mathbf{W}}^{\boldsymbol{\theta}})+\sum_{l=1}^{K_j}\overline{\mathbf{W}}_{jl}^{\boldsymbol{\theta}}(y)e_{jl}(y)\}dv(y)$, $\mathcal{E}_{ij}(y;\mathbf{F})=\sum_{\delta_{ik}=0}\mathbf{F}_{ik}(y)[\int_{X_{ik}}^{\tau}g_{\mathrm{D}}\{\boldsymbol{\eta}_{\boldsymbol{\alpha}_0}(u,s_{ik})+\boldsymbol{\beta}_0'\mathbf{Z}_i+h_0(y)\}d\mathfrak{s}_j(u+s_{ik},X_i;\mathbf{Z}_i)+g_{\bar{\mathrm{D}}}\{\xi_{\boldsymbol{a}_0}(s_{ik})+\mathbf{b}_0'\mathbf{Z}_i+c_0(y)\}\mathfrak{s}_j(\tau,X_i;\mathbf{Z}_i)]$.

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

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

where

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

It follows that

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

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). Semiparametric receiver operating characteristic analysis to evaluate biomarkers for disease. *J. Am. Statist. Assoc.* **97**, 1099–1107.

CAI, T., PEPE, M. S., LUMLEY, T., ZHENG, Y. & JENNEY, N. S. (2003). The sensitivity and specificity of markers for event times. *UW Biostatistics Working Paper Series* .

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

CARNEY, P., MIGLIORETTI, D., YANKASKAS, B., KERLIKOWSKE, K., ROSENBERG, R. & RUTTER, C. (2003). Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening. *Mammography* **138**, 168–75.

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.

DE GRUTTOLA, V. & TU, X. M. (1994). Modelling progression of Cd4-lymphocyte count and its relationship to survival time. *Biometrics* **50**, 1003–1014.

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., PEARSON, 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). *The Statistical Evaluation of Medical Tests for Classification and Prediction.* 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.

ROTNITZKY, A. & ROBINS, J. M. (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika* **82**, 805–820.

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: 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

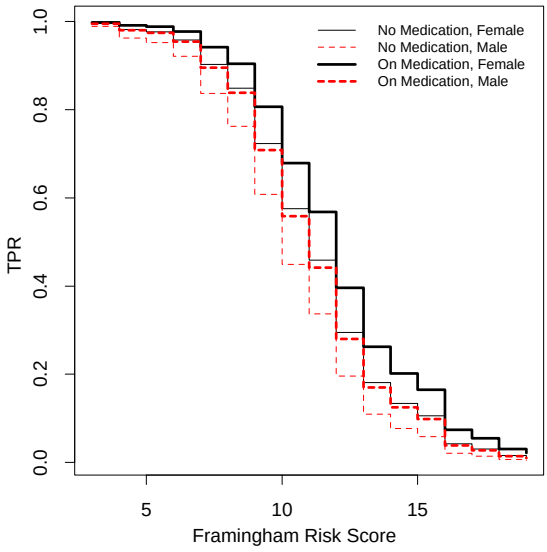
Table 2: Estimated AUCs and their standard errors. Shown also are the differences in AUCs between groups and between years. Δ_{Year}, Δ_{Gender} and Δ_{Med} denote the difference in AUC between Year 1 and Year 5, between Female and Male, and between No medication and with medication, respectively.

	Year 1			Year 5			Δ_{Year}	
	Female	Male	Δ_{Gender}	Female	Male	Δ_{Gender}	Female	Male
No Med.	.691(.026)	.618(.022)	.073(.022)	.617(.023)	.541(.019)	.076(.023)	.074(.023)	.077(.024)
Yes Med.	.666(.027)	.589(.026)	.077(.024)	.587(.027)	.508(.025)	.079(.025)	.078(.025)	.081(.025)
Δ_{Med}	.026(.023)	.030(.024)	–	.030(.024)	.033(.024)	–		

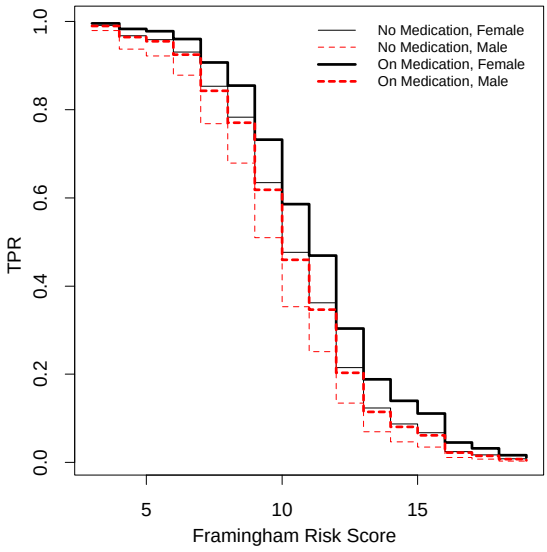
Table 3: Estimated AUCs for the cumulative incidence based ROC curves and their standard errors. Shown also are the differences in AUCs between groups and between years. Δ_{Year}, Δ_{Gender} and Δ_{Med} denote the difference in AUC between Year 1 and Year 5, between Female and Male, and between No medication and with medication, respectively.

	Year 1			Year 5			Δ_{Year}	
	Female	Male	Δ_{Gender}	Female	Male	Δ_{Gender}	Female	Male
No Med.	.701(.028)	.635(.025)	.066(.019)	.634(.022)	.561(.016)	.073(.022)	.068(.020)	.075(.021)
Yes Med.	.677(.029)	.611(.027)	.067(.020)	.606(.024)	.530(.021)	.075(.023)	.072(.022)	.080(.022)
Δ_{Med}	.024(.020)	.025(.019)	–	.028(.023)	.031(.023)	–		

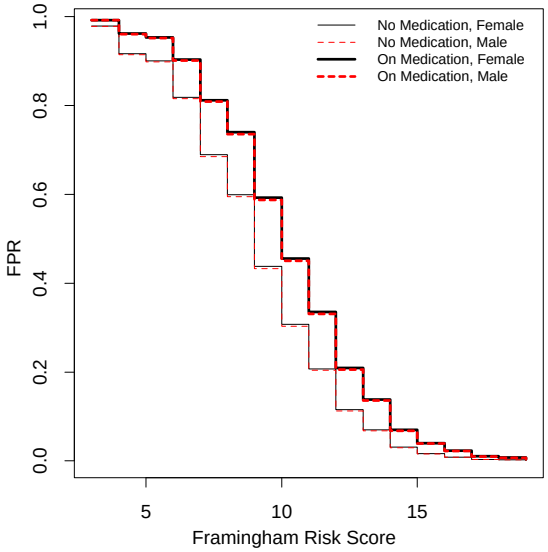
Figure 1: Estimated TPR / FPR functions of the Framingham risk score for all 4 groups: female subjects (solid curves) / male subjects (dashed curves) who are on medication (thicker curves) / not on medication (thinner curves).



(a) TPR : $T = 1$ year

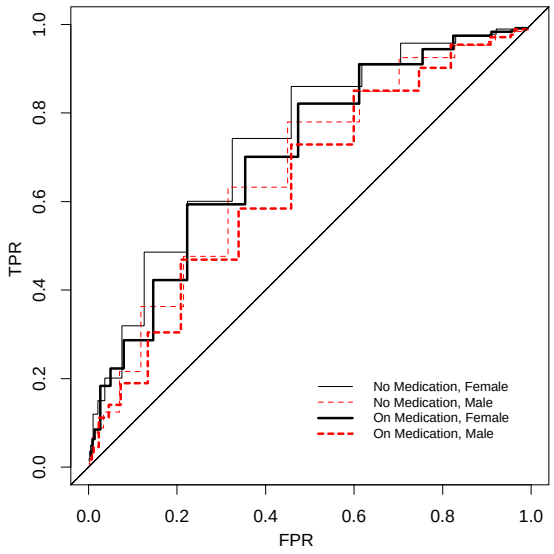


(b) TPR : $T = 5$ year

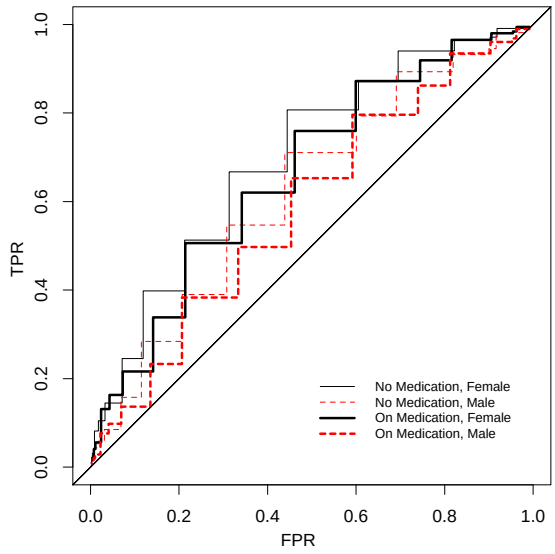


(c) FPR : $T > \tau = 7$ years

Figure 2: 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.

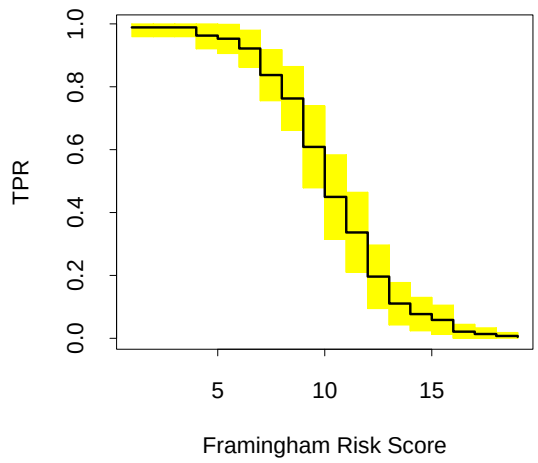


(a) $t = 1$ year

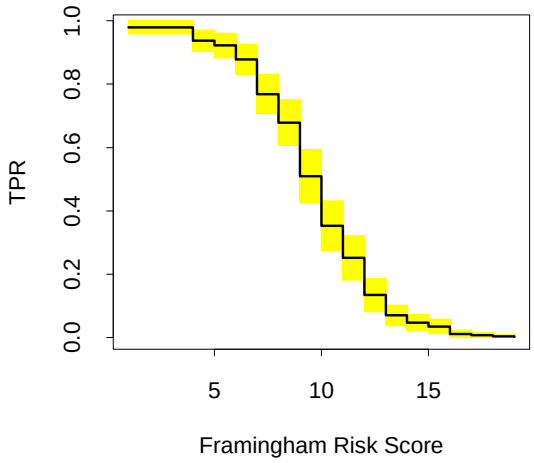


(b) $t = 5$ years

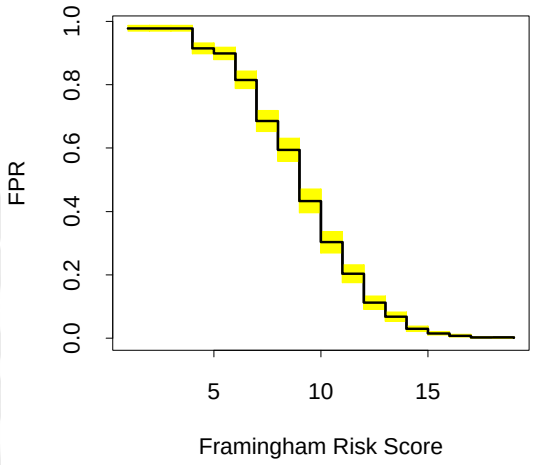
Figure 3: Estimated TPR / FPR functions of the Framingham risk score for male subjects who are not on medication. The sensitivities (TPR) for events at $t = 1$ and $t = 5$ years after the FR-score is measured are displayed. Shown also are their 95% simultaneous confidence bands (shaded regions).



(a) TPR : $T = 1$ year

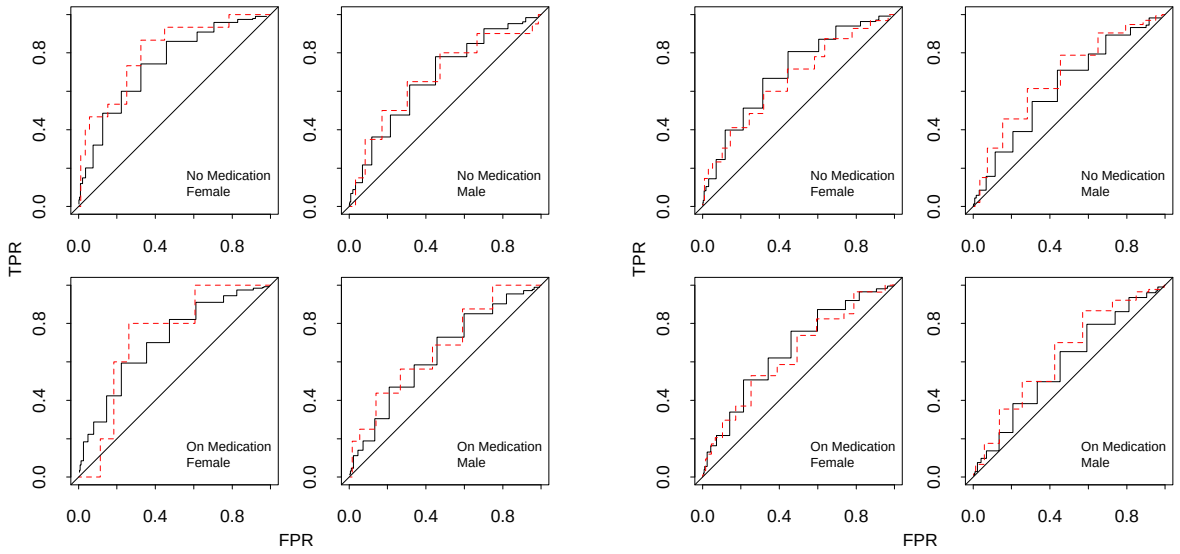


(b) TPR : $T = 5$ year



(c) FPR : $T > \tau = 7$ years

Figure 4: Plots of the cumulative incidence based ROC curve: at $t = 1$ year and $t = 5$ years. Shown are the semi-parametric model based estimates (solid curves) and the non-parametric estimates (Heagerty et al, 2000) (dashed curves).



(a) $t = 1$ year

(b) $t = 5$ years