11-14-2007

A Parametric ROC Model Based Approach for Evaluating the Predictiveness of Continuous Markers in Case-control Studies

Ying Huang
University of Washington, ying@u.washington.edu

Margaret Pepe
University of Washington, Fred Hutch Cancer Research Center, mspepe@u.washington.edu

1. Background

The importance of biomarkers in disease screening, diagnosis, and risk prediction has been generally recognized. A well-established criterion for biomarker selection is classification accuracy, commonly characterized by the Receiver Operating Characteristic (ROC) curve and its summary measures. However, classification is not always the major focus. Oftentimes we use biomarkers to calculate the risk of an outcome. Recently, there has been increasing awareness that the ROC curve is not the most relevant tool for assessing a biomarker whose purpose is risk prediction (Gail and Pfeiffer, 2005; Huang et al., 2007; Pepe et al., 2007; Cook, 2007). On the one hand, the ROC curve does not display risk which is of primary interest to patients and clinicians. On the other hand, criteria relating to classification oftentimes can be too stringent for evaluation of a risk prediction marker. To characterize the predictive capacity of a continuous marker or risk model, Huang et al. (2007) proposed a new graphical tool, the predictiveness curve, to display the population distribution of disease risk predicted by the particular marker or risk model.

Let D denote a binary outcome that we term disease here, $D = 1$ for diseased and $D = 0$ for non-diseased. Let Y denote a vector of predictors of interest and $Risk(Y) = P(D = 1|Y)$ denote the risk calculated on the basis of Y. The predictiveness curve for Y is the curve $R(v)$ vs v for $v \in (0, 1)$, where $R(v)$ is the vth percentile of $Risk$. The inverse function

$$R^{-1}(p) = P(Risk \leq p),$$

is the proportion of the population with risks less than or equal to p. In other words $R^{-1}(p)$ is the population cumulative distribution function of risk. An appealing property of the predictiveness curve is that it provides a common meaningful scale for making comparisons between markers or risk models that may not be comparable on their original scales. Comparisons might be based on $R(v)$, the risk percentiles. A better risk prediction marker tends to have larger variability in $R(v)$. A clinically compelling comparison is based on $R^{-1}(p)$. Suppose there exists a low risk threshold p_L and/or a high risk threshold p_H which are agreed upon apriori such that the decision for treatment

1

is recommendation for or against if the estimated risk for a patient is above p_H or below p_L. A marker or risk model is preferable to another if it categorizes more people into the low and high risk ranges where decisions are easy to make and leaves fewer subjects in the equivocal risk range. That is, we hope to identify markers that have large values of $R^{-1}(p_L)$ and $1 - R^{-1}(p_H)$ and small values of $R^{-1}(p_H) - R^{-1}(p_L)$.

Figure 1 displays ROC and predictiveness curves for PSA and PSA velocity as classification or risk prediction markers for prostate cancer. Details of the data will be discussed in Section 4. The ROC curves shown in Figure 1(a) suggest that PSA has better classification accuracy than PSA velocity. The corresponding predictiveness curves are shown in Figure 1(b). We see for example that at the 90th percentile the risk is 0.372 for PSA but only 0.295 for PSA velocity suggesting PSA is a better marker of high risk than PSA velocity. PSA is also a better marker of low risk. The 10th percentile of risk is 0.145 according to PSA velocity but much lower based on PSA, 0.091. We can also consider the inverse function taking $p_H = 0.29$ and $p_L = 0.10$. PSA is predictive of low risk in $R^{-1}(0.10) = 13.6\%$ of the population, more than 3.0%, the proportion identified with PSA velocity. PSA also identifies more people at high risk than does PSA velocity with $1 - R^{-1}(0.29) = 25.5\%$ and 11.1% respectively. Less patients are categorized into the equivocal risk range according to PSA (60.9%) than PSA velocity (85.9%).

Semi-parametric estimators for making inference about the curve and for making pointwise comparisons between two curves from a cohort study have been developed by Huang et al. (2007). Since case-control studies are often performed in the early phases of biomarker development (Pepe et al., 2001), it is important to estimate the predictiveness of continuous biomarkers in studies that use this type of design as well. When the disease of interest is rare in the population, using a case-control sampling to oversample cases can be more efficient than simple random sampling from the population. In this paper, we consider estimation of the predictiveness curve from a case-control study based on modeling a parametric ROC curve, assuming the disease prevalence is known apriori. The idea behind this is the one-to-one relationship between the predictiveness
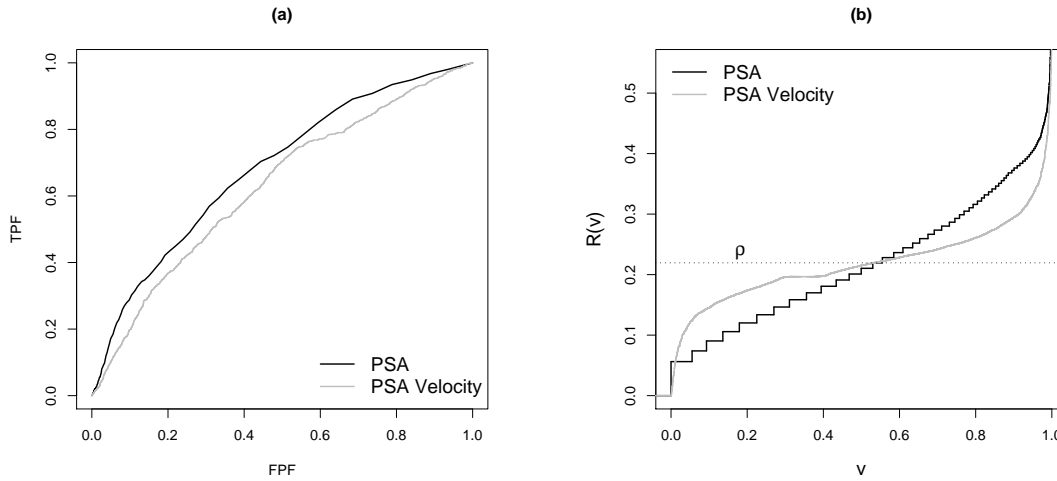
2

Figure 1. ROC curves and Predictiveness curves for two markers of prostate cancer estimated from the PCPT cohort data.

curve and the ROC curve.

2. Relationship between the Predictiveness Curve and the ROC Curve

Here we focus on the scenario of a single continuous marker. Note the marker may be a predefined combination of predictors such as the Framingham risk score. Denote Y, Y_D, and $Y_{\bar{D}}$ as the marker measurement in the general, diseased, and non-diseased populations respectively. Let F, F_D, and $F_{\bar{D}}$ be the corresponding distribution functions and let f, f_D, and $f_{\bar{D}}$ be the density functions. We assume the disease prevalence $\rho = P(D = 1)$ is known and that the risk of disease $P(D = 1|Y)$ is monotone increasing in Y.

Under this monotone increasing risk assumption, we have $R(v) = P\{D = 1|Y = F^{-1}(v)\}$. The following theorem characterizes the one-to-one relationship between the predictiveness curve and the ROC curve.

Theorem 1

Suppose Y_D and $Y_{\bar{D}}$ have absolutely continuous distribution functions and $P(D = 1|Y)$ is monotone increasing in Y. Further suppose the support of $Y_{\bar{D}}$ covers the support of Y_D. Then

3

$R(v)$ vs v can be represented as

$$\frac{\rho \text{ROC}'(t)}{\rho \text{ROC}'(t) + (1-\rho)} \quad \text{vs} \quad 1 - (1-\rho)t - \rho \text{ROC}(t), \qquad t \in (0,1), \tag{1}$$

where the ROC curve at false positive rate t is $\text{ROC}(t)$, and $\text{ROC}'(t)$ is its derivative with respect to t.

Proof

For $v \in (0,1)$, let $y = F^{-1}(v)$. Suppose $F_{\bar{D}}(y) = 1 - t$, since y is within the support of $Y_{\bar{D}}$, we have $y = F_{\bar{D}}^{-1}(1-t)$ trivially. Let $\mathcal{L}\text{R}$ denote the likelihood ratio function: $\mathcal{L}\text{R}(y) = f_D(y)/f_{\bar{D}}(y)$. We have

$$
\begin{aligned}
v = F(y) &= (1-\rho)F_{\bar{D}}(y) + \rho F_D(y) \\
&= (1-\rho)(1-t) + \rho F_D(F_{\bar{D}}^{-1}(1-t)) \\
&= (1-\rho)(1-t) + \rho \{1 - \text{ROC}(t)\} \\
&= 1 - (1-\rho)t - \rho \text{ROC}(t).
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
R(v) &= P\{D = 1 | Y = y\} \\
&= \frac{\rho \mathcal{L}\text{R}(y)}{\rho \mathcal{L}\text{R}(y) + (1-\rho)} \\
&= \frac{\rho \mathcal{L}\text{R}\{F_{\bar{D}}^{-1}(1-t)\}}{\rho \mathcal{L}\text{R}\{F_{\bar{D}}^{-1}(1-t)\} + (1-\rho)} \\
&= \frac{\rho \text{ROC}'(t)}{\rho \text{ROC}'(t) + (1-\rho)}.
\end{aligned}
$$

The last equality holds since $\mathcal{L}\text{R}\{F_{\bar{D}}^{-1}(1-t)\} = \text{ROC}'(t)$ (Pepe, 2003). Note the result can be generalized to the scenario when the upper bound of the support for Y_D is larger than the upper bound of the support for $Y_{\bar{D}}$. We omit the details because it is not relevant for the method discussed

4

in this paper. ∎

Theorem 1 has implications for estimating the predictiveness curve from a case-control sample with n_D cases and $n_{\bar{D}}$ controls. We can estimate a smooth ROC curve first and then derive the corresponding predictiveness curve based on (1). This is an appealing procedure for the following reasons: (i) The fact that the ROC curve can be estimated only from ranked data implies that methods for deriving the predictiveness curve from a rank-based ROC curve estimate also only depend on ranks. This contrasts with previous methods proposed (Huang et al., 2007). (ii) Estimation of the ROC curve is a well studied problem. There are a wide variety of methods available. (iii) It is natural to estimate the ROC curve from a case-control study since sensitivity and specificity are defined conditional on disease status. (iv) The area under the estimated predictiveness curve that is derived from an estimated ROC curve is always equal to ρ, as shown below. This is a fundamental property of the predictiveness curve since $\int_0^1 R(v)dv = E(Risk) = E(D) = \rho$.

$$
\begin{aligned}
& \int_0^1 R(v)dv \\
=\ & \int_{t=1}^{t=0} \frac{\rho\text{ROC}'(t)}{\rho\text{ROC}'(t) + (1-\rho)} d\left\{1 - (1-\rho)t - \rho\text{ROC}(t)\right\} \\
=\ & -\int_{t=1}^{t=0} \frac{\rho\text{ROC}'(t)}{\rho\text{ROC}'(t) + (1-\rho)} \left\{(1-\rho) + \rho\text{ROC}'(t)\right\} dt \\
=\ & \rho \int_{t=0}^{t=1} \text{ROC}'(t)dt \\
=\ & \rho\left\{\text{ROC}(1) - \text{ROC}(0)\right\} \\
=\ & \rho
\end{aligned}
\tag{2}
$$

This result holds for the ROC curve based on a continuous marker as well as the ROC curve constructed from a discrete marker which has finite derivative everywhere (details omitted). When we compare two markers with respect to the steepness of their estimated predictiveness curves, it facilitates visual comparisons when the estimated curves have the same area under the curve.

Approaches to estimate an ROC curve vary along a spectrum regarding assumptions made. At

5

the one extreme, we can model the marker distributions within cases and controls parametrically and estimate the corresponding ROC curve. For example Wieand et al. (1989) modeled Y_D and $Y_{\bar{D}}$ as normally distributed. At the other extreme, an ROC curve can be estimated completely nonparametrically using empirical estimators for F_D and $F_{\bar{D}}$ (Greenhouse and Mantel, 1950; Hsieh and Turnbull, 1996). A method in-between is to assume a parametric model for the ROC curve without enforcing any parametric distributional assumptions on marker measures. This semi-parametric approach is more efficient than the nonparametric approach, yet more robust than modeling the marker distributions parametrically. A review of these approaches will be given later.

Before we start exploring different parametric ROC models, we note that the assumption that $P(D = 1|Y = y)$ is increasing in y implies that $\mathcal{L}R(y)$ is increasing in y which in turn implies that $\text{ROC}'(t)$ is decreasing in t. That is, a monotone increasing risk assumption implies concavity of the corresponding ROC curve. Therefore, we prefer methods that lead to concave estimates of the ROC curves. Note that concavity has always been a desirable property for an ROC curve because it guarantees that the ROC curve will never cross the 45° "guessing line" (Dorfman et al., 1996) and because it is a property of the optimal ROC curve for decision rules based on y (McIntosh and Pepe, 2002).

3. Parametric ROC Models

3.1 *The Binormal ROC Curve*

The most widely used parametric ROC model is the binormal ROC curve. It assumes that there exists a common monotone transformation h, which transforms the marker distributions in both cases and controls to normality. Suppose $h(Y_{\bar{D}}) \sim N(\mu_{\bar{D}}, \sigma_{\bar{D}}^2)$, $h(Y_D) \sim N(\mu_D, \sigma_D^2)$, the corresponding ROC curve is

$$\text{ROC}(t) = \Phi\{a + b\Phi^{-1}(t)\},$$

where $a = (\mu_D - \mu_{\bar{D}})/\sigma_D$ and $b = \sigma_{\bar{D}}/\sigma_D$. Many algorithms have been proposed to fit the binormal ROC curve. Moreover, the binormal assumption is thought to fit many real datasets (Hanley, 1988).

6

However, a problem with using the binormal ROC model is that it is not concave in $(0, 1)$ unless $b = 1$ (i.e. the normal distributions for cases and controls have the same variance). This can be recognized from the quadratic form of the derivative of the binormal ROC curve,

$$\text{ROC}'(t) = \frac{b\phi\{a + b\Phi^{-1}(t)\}}{\phi\{\Phi^{-1}(t)\}} = C^+ \exp\left[-\frac{1}{2}(b^2 - 1)\left\{\Phi^{-1}(t) + \frac{ab}{b^2 - 1}\right\}^2\right], \tag{3}$$

where C^+ is some positive constant.

Thus, if we assume the risk function is monotone increasing as Y increases, the binormal model is not a suitable model for the ROC curve unless we assume $\sigma_D = \sigma_{\bar{D}}$.

3.2 *Concave Parametric ROC Models*

Two published parametric models for concave ROC curves are the bigamma and bilomax models. The bigamma ROC curve (Dorfman et al., 1996) assumes there exists a common monotone transformation that transforms the distributions of Y_D and $Y_{\bar{D}}$ into gamma distributions with the same shape parameter, an extension to the one-parameter exponential family of England (1988). The use of this ROC model is hindered by the fact that the ROC function cannot be written down in closed-form. The bilomax ROC curve proposed by Campbell and Ratnaparkhi (1993) assumes the existence of a monotone transformation h such that the distributions of $h(Y_{\bar{D}})$ and $h(Y_D)$ are lomax or type II Pareto (Lomax, 1954). That is,

$$f_{\bar{D}}\{h(Y_{\bar{D}})\} = \frac{b_1 k_1}{\{1 + b_1 h(Y_{\bar{D}})\}^{k_1+1}}, \qquad b_1 > 0, k_1 > 0,$$
$$f_D\{h(Y_D)\} = \frac{b_2 k_2}{\{1 + b_2 h(Y_D)\}^{k_2+1}}, \qquad b_2 > 0, k_2 > 0.$$

where b_1, b_2 and k_1, k_2 are scale and shape parameters respectively. The corresponding ROC curve has an explicit form:

$$\text{ROC}(t) = \left\{1 + \frac{b_2}{b_1}\left(t^{-1/k_1} - 1\right)\right\}^{-k_2} \equiv \left\{1 + b\left(t^{-1/k_1} - 1\right)\right\}^{-k_2}.$$

7

This three-parameter bilomax ROC curve is concave if $b = b_2/b_1 < 1$ and $k_2 \leq k_1$.

For estimation of a whole predictiveness curve, we choose estimation of the bilomax ROC curve as an intermediate step. In practice, the likelihood based on the three-parameter model can be very flat leading to numerical difficulties in maximization (Campbell and Ratnaparkhi, 1993). Hence we use a two-parameter bilomax ROC model that restricts $k_1 = k_2 = k$. This simplifies numerical fitting procedures considerably with only a minor loss in flexibility. Under the two-parameter bilomax ROC model $\text{ROC}(t) = \left\{ 1 + b \left(t^{-1/k} - 1 \right) \right\}^{-k}$, we have $\text{ROC}'(t) = bt^{-1/k-1} \left\{ 1 + b(t^{-1/k-1}) \right\}^{-k-1}$.

3.3 Estimation

Denote by $\text{ROC}_\theta(t)$, $t \in (0,1)$, the parametric ROC curve with parameter θ. Define

$$G(\theta, t) = \frac{\rho \text{ROC}'_\theta(t)}{\rho \text{ROC}'_\theta(t) + (1 - \rho)}$$

and

$$H(\theta, t) = 1 - (1 - \rho)t - \rho \text{ROC}_\theta(t).$$
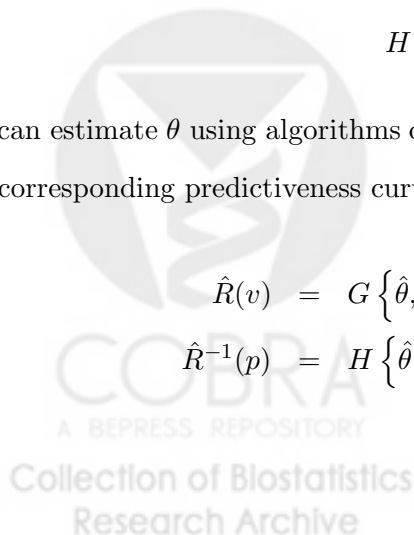
Let

$$G^{-1}(\theta, p) = \inf \left\{ t : G(\theta, t) \leq p \right\}$$

and

$$H^{-1}(\theta, v) = \left\{ t : H(\theta, t) = v \right\}.$$

We can estimate θ using algorithms described below and denote its estimation by $\hat{\theta}$. Estimators of the corresponding predictiveness curve and its inverse are

$$
\begin{aligned}
\hat{R}(v) &= G \left\{ \hat{\theta}, H^{-1}(\hat{\theta}, v) \right\} && \text{for } v \in (0,1), \\
\hat{R}^{-1}(p) &= H \left\{ \hat{\theta}, G^{-1}(\hat{\theta}, p) \right\} && \text{for } p \in \left\{ R(v) : v \in (0,1) \right\}.
\end{aligned}
$$

8

In practice, oftentimes there does not exist a closed-form for $H^{-1}(\cdot)$ or $G^{-1}(\cdot)$ and numerical methods need to be implemented to calculate $\hat{R}(v)$ and $\hat{R}^{-1}(p)$.

There are many existing semi-parametric approaches we can use to estimate θ for a whole ROC curve. Metz et al. (1998) proposed grouping continuous data and estimating the parameters based on the Dorfman and Alf maximum-likelihood algorithm for ordinal data (Dorfman and Alf, 1969). Hsieh and Turnbull (1996) developed a generalized least squares method to fit a parametric ROC curve to discretized continuous data. Pepe (2000) and Alonzo and Pepe (2002) proposed a distribution-free ROC-GLM procedure. Zou and Hall (2000) developed maximum likelihood rank-based estimator as solution to the score equations derived from the likelihood function of the order statistics. Pepe and Cai (2004) maximized the pseudolikelihood based on the standardized marker value. Cai and Moskowitz (2004) developed a maximum profile likelihood approach which provides a fully efficient parameter estimate. These semi-parametric approaches have the advantage of being rank-based. We can also model the marker distributions parametrically (Wieand et al., 1989), the corresponding ROC curve and the predictiveness curve estimators, however, do not have this rank-invariance property.

3.4 *Modeling a Portion of the Predictiveness Curve*

3.4.1 Boundary Problems   In fitting a parametric risk model to data, we assume the model holds over the observed range of the data. In contrast, for the parametric-ROC-model, the parametric form of $\mathrm{ROC}'_\theta(t)$ is assumed to hold as $t \to 0$ and $t \to 1$. The heavy reliance of $\lim_{t\to 0} \mathrm{ROC}'_\theta(t)$ and $\lim_{t\to 1} \mathrm{ROC}'_\theta(t)$ on the parametric model causes a lack of flexibility in the estimated predictiveness curve, especially at the boundary. For example, with a binormal ROC model, $\lim_{t\to 1} \mathrm{ROC}'(t) = 0$ or ∞, which implies $Risk(y) = 0$ or 1 as $y \to \infty$ and consequently $R(1) = 0$ or 1, which may not be reasonable in practice.

As shown in Table 1, when a binormal ROC model is enforced, the boundary of the predictiveness curve has value either 0 or 1, completely determined by the parameter b, the ratio of standard deviation between the latent case and control marker distributions. When a bilomax ROC model

9

Table 1

Properties of parametric ROC models and the corresponding predictiveness curves.

$\lim \mathrm{ROC}'(t)$	Implications for $\lim R(v)$
Binormal $\mathrm{ROC}(t) = \Phi\{a + b\Phi^{-1}(t)\}$	
$b < 1$	
$\lim_{t \to 1} \mathrm{ROC}'(t) \to \infty$	$\lim_{v \to 0} R(v) = 1$
$\lim_{t \to 0} \mathrm{ROC}'(t) \to \infty$	$\lim_{v \to 1} R(v) = 1$
$b > 1$	
$\lim_{t \to 1} \mathrm{ROC}'(t) \to 0$	$\lim_{v \to 0} R(v) = 0$
$\lim_{t \to 0} \mathrm{ROC}'(t) \to 0$	$\lim_{v \to 1} R(v) = 0$
$b = 1$	
$\lim_{t \to 1} \mathrm{ROC}'(t) \to 0$	$\lim_{v \to 0} R(v) = 0$
$\lim_{t \to 0} \mathrm{ROC}'(t) \to \infty$	$\lim_{v \to 1} R(v) = 1$
Bilomax $\mathrm{ROC}(t) = \left\{1 + b\left(t^{-1/k_1} - 1\right)\right\}^{-k_2}$	
$k_2 < k_1$	
$\lim_{t \to 1} \mathrm{ROC}'(t) \to bk_2/k_1$	$\lim_{v \to 0} R(v) = \frac{\rho}{\rho + (1-\rho)k_1/(bk_2)}$
$\lim_{t \to 0} \mathrm{ROC}'(t) \to \infty$	$\lim_{v \to 1} R(v) = 1$
$k_2 = k_1 = k$	
$\lim_{t \to 1} \mathrm{ROC}'(t) \to b$	$\lim_{v \to 0} R(v) = \frac{\rho}{\rho + (1-\rho)/b}$
$\lim_{t \to 0} \mathrm{ROC}'(t) \to 1/b^k$	$\lim_{v \to 1} R(v) = \frac{\rho}{\rho + (1-\rho)b^k}$

is assumed, there is more flexibility compared to a binormal ROC model at the boundary of the predictiveness curve estimate. But still the values of $\hat{R}(v)$ as v approaches 0 and 1 depend entirely on the parameter estimates of the corresponding ROC curve.

We propose to modify the approach described in Section 3.3 such that the predictiveness curve estimator is more data-dependent at the boundary without losing the rank-invariance property. In particular we consider modelling a portion of the ROC curve.

3.4.2 The Partial Predictiveness Curve   Heretofore our interest was in estimating the predictiveness curve over the whole domain $v \in (0,1)$ by first estimating a parametric ROC model over its whole domain $t \in (0,1)$. However, to evaluate a marker for its risk prediction capability, sometimes we do not need the whole predictiveness curve if only risks within a particular range are of primary interest. We can instead examine a segment of the predictiveness curve covering those points of

10

major concern. To derive this "partial predictiveness curve" using a ROC curve based strategy, only a portion of the ROC curve needs to be modeled.

Researchers in the field of diagnostic test evaluation have long been interested in the partial ROC curve. For example, in screening studies, it is important to minimize the unnecessary cost due to false positive test results, hence the region of the ROC curve corresponding to low FPF is most relevant. If the purpose of the study is disease diagnosis, it is critical not to miss detecting subjects with disease, and hence the part of the ROC curve corresponding to high TPF is of primary interest. Modeling a partial ROC curve and the area under it has been proposed and studied (McClish, 1989; Thompson and Zucchini, 1989; Jiang et al., 1996; Dodd and Pepe, 2003; Pepe and Cai, 2004).

Interestingly, when concavity is required only over a certain portion of the ROC curve, parametric ROC models which are not concave in the whole range may be employed. Consider the classic binormal ROC curve, $\text{ROC}(t) = \Phi\left\{a + b\Phi^{-1}(t)\right\}$, whose derivative is shown in (3). Consider the following two scenarios.

(i) If $0 < b < 1$, $\text{ROC}'(t)$ increases as $\left\{\Phi^{-1}(t) + ab/(b^2 - 1)\right\}^2$ increases. Thus for $\text{ROC}'(t)$ to be monotone decreasing, we need to have

$$\Phi^{-1}(t) < \frac{-ab}{b^2 - 1} \iff t < \Phi\left(\frac{-ab}{b^2 - 1}\right) \iff a > \frac{1 - b^2}{b}\Phi^{-1}(t).$$

That is, a portion of the ROC curve with small FPF could be modeled to be concave.

(ii) If $b > 1$, $\text{ROC}'(t)$ increases as $\left\{\Phi^{-1}(t) + ab/(b^2 - 1)\right\}^2$ decreases. Thus for $\text{ROC}'(t)$ to be monotone decreasing, we need to have

$$\Phi^{-1}(t) > \frac{-ab}{b^2 - 1} \iff t > \Phi\left(\frac{-ab}{b^2 - 1}\right) \iff a > \frac{1 - b^2}{b}\Phi^{-1}(t).$$

That is, a portion of the ROC curve with large TPF could be modeled to be concave.

We can impose this restriction during estimation to guarantee concavity of the partial binormal ROC curve over the range of interest. Suppose concavity is required for $t \in (t_0, t_1)$, we can fit the

11

ROC model with the restriction on parameter estimates (a, b):

$$a > \frac{1 - b^2}{b} \Phi^{-1}(t_0) \quad \& \quad a > \frac{1 - b^2}{b} \Phi^{-1}(t_1).$$

3.5 *Asymptotic Theory for the Estimator of the Predictiveness Curve Derived from a Parametric ROC Model*

Suppose $\hat{R}(v)$ and $\hat{R}^{-1}(p)$ are estimators of $R(v)$ and $R^{-1}(p)$ based on either modeling a whole range or a portion of the predictiveness curve. We assume the following conditions hold for asymptotic theory.

Assumptions

(i) $\sqrt{n} \left(\hat{\theta} - \theta \right) \xrightarrow{d} N(0, \Sigma(\theta))$.

(ii) $\mathrm{ROC}_\theta(t)$ is differentiable with respect to θ and t.

(iii) $G(\theta, t)$ is differentiable with respect to θ and t with derivatives g_1 and g_2.

(iv) $H^{-1}(\theta, v)$ is differentiable with respect to θ.

(v) $G^{-1}(\theta, p)$ is differentiable with respect to θ.

Theorem 2

$$\sqrt{n} \left\{ \hat{R}(v) - R(v) \right\} \xrightarrow{d} N\left(0, \sigma^2(v)\right),$$

where

$$
\begin{aligned}
\sigma^2(v) \quad = \quad & \left[g_1\left\{\theta, H^{-1}(\theta, v)\right\} + g_2\left\{\theta, H^{-1}(\theta, v)\right\} \left(\frac{\partial H^{-1}(\theta, v)}{\partial \theta} \right) \right]^T \Sigma(\theta) \\
& \left[g_1\left\{\theta, H^{-1}(\theta, v)\right\} + g_2\left\{\theta, H^{-1}(\theta, v)\right\} \left(\frac{\partial H^{-1}(\theta, v)}{\partial \theta} \right) \right].
\end{aligned}
$$

∎

Theorem 3

$$\sqrt{n} \left\{ \hat{R}^{-1}(p) - R^{-1}(p) \right\} \xrightarrow{d} N\left(0, \tau^2(p)\right),$$

where

$$\tau^2(p) \;\; = \;\; \left(\frac{\partial H \left\{ \theta, G^{-1}(\theta, p) \right\}}{\partial \theta} \right)^T \Sigma(\theta) \left(\frac{\partial H \left\{ \theta, G^{-1}(\theta, p) \right\}}{\partial \theta} \right),$$

where

$$\frac{\partial H \left\{ \theta, G^{-1}(\theta, p) \right\}}{\partial \theta} = \rho \frac{\partial \mathrm{ROC}_\theta \left\{ G^{-1}(\theta, p) \right\}}{\partial \theta} + \left[1 - \rho + \rho \mathrm{ROC}'_\theta \{ G^{-1}(\theta, p) \} \right] \frac{\partial G^{-1}(\theta, p)}{\partial \theta}.$$

∎

Observe that $\sigma^2(v) = \{ \partial R(v) / \partial v \}^2 \tau^2(p) = \left[g_2 \left\{ \theta, H^{-1}(\theta, v) \right\} \partial H^{-1}(\theta, v) / \partial v \right]^2 \tau^2(p)$ for $p = R(v)$. Theorems 2 and 3 follow directly from the continuous mapping theorem and chain rule. In practice, due to the lack of closed-forms for $H^{-1}(\cdot)$ and $G^{-1}(\cdot)$, numerical differentiation methods are needed for calculation of their derivatives, in order to estimate the variances of $\hat{R}(v)$ and $\hat{R}^{-1}(p)$.

4. Markers for Prostate Cancer

The Prostate Cancer Prevention Trial (PCPT) was a randomized prospective study of men with PSA < 3.0 ng/mL and aged 55+ years who were followed up for 7 years with annual PSA measurements. A biopsy was recommended for all men either during or at the end of the study. Thompson et al. (2006) identified 5519 men on the placebo arm of the trial who had undergone prostate biopsy and had a PSA and digital rectal exam (DRE) during the year prior to biopsy and at least 2 PSA values from the 3 years prior to biopsy. The risk of finding prostate cancer in the biopsy was evaluated as a function of PSA, PSA velocity and several other variables including age, family history, DRE and prior prostate biopsy. The concept is that in the future a man may use his calculated risk to decide whether or not to have a biopsy procedure performed. We use this study cohort as our population of interest and illustrate estimation of the predictiveness curve using a case-control subset from the cohort. 250 cases and 250 controls are randomly sampled, from which we compare PSA and PSA velocity as predictors of prostate cancer risk. 21.9% of men were found to have prostate cancer in the original cohort. This number is treated as the "known" prevalence in our
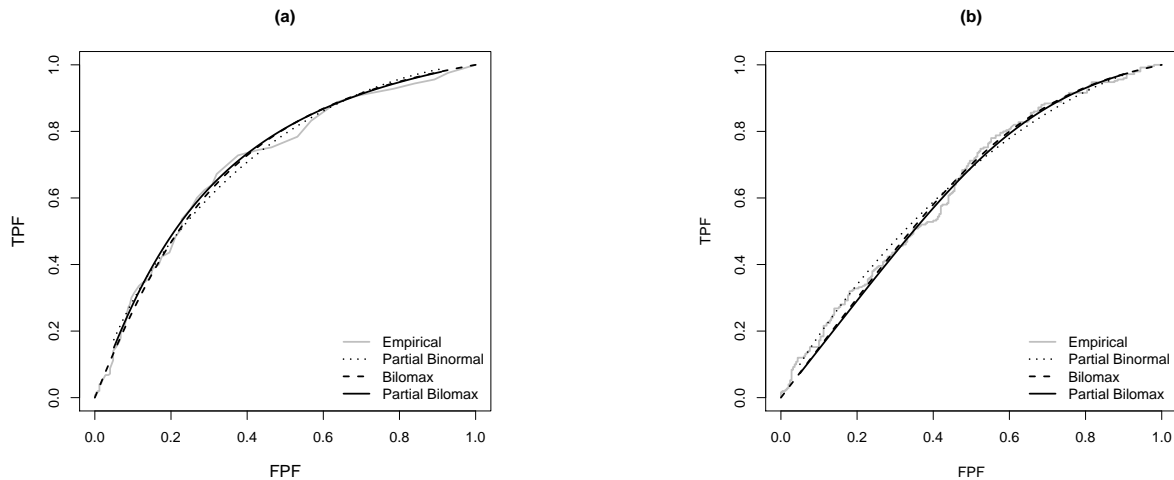
13

Figure 2. ROC curve estimates from the PCPT data: (a) PSA, (b) PSA velocity.

analysis.

We fit a two-parameter bilomax ROC curve, a partial bilomax ROC curve with $t \in (0.05, 0.91)$, and a partial binormal ROC curve with $t \in (0.05, 0.91)$ to the case-control data, from which the predictiveness curves are derived. We modified the maximum profile likelihood method for fitting a binormal ROC curve (Cai and Moskowitz, 2004) to fit the bilomax ROC curve. The placement-value-based pseudolikelihood approach (Pepe and Cai, 2004) is used to fit the partial bilomax and binormal ROC curves. The estimated ROC curves for PSA and PSA velocity are displayed in Figure 2. Also displayed are the empirical ROC curves (Obuchowski, 2003) estimated from the case-control sample. The parametric ROC curve estimates in general agree fairly well with the empirical ROC curves. The difference between the whole and partial bilomax ROC curves appears to be smaller compared to the differences between them and the partial binormal ROC curve. PSA appears to have much better classification accuracy than PSA velocity for diagnosing prostate cancer.

Figure 3 displays the corresponding predictiveness curves estimated from the case-control sample for PSA and PSA velocity as well as the empirical predictiveness curve estimate. The empirical

14

predictiveness curve is generated in the following way. The observations are partitioned into 10 groups according to cutpoints y_1, y_2, ..., y_9. Let Y_{Dj} be the jth case and let $Y_{\bar{D}i}$ be the ith control. Let $y_{10} = 1$, for $k \in \{1, \ldots, 10\}$, we calculate

$$v_k = \rho \frac{1}{n_D} \sum_{j=1}^{n_D} I(Y_{Dj} \leq y_k) + (1 - \rho) \frac{1}{n_{\bar{D}}} \sum_{i=1}^{n_{\bar{D}}} I(Y_{\bar{D}i} \leq y_k).$$

We then calculate $P(D = 1|k, \text{ sampled})$, the average risk within the kth group in the case-control sample, using sample mean of D. The population risk within k^{th} group, $P(D = 1|k)$ is calculated based on the Bayes theorem

$$\frac{P(D = 1|k)}{1 - P(D = 1|k)} = \frac{P(D = 1|k, \text{ sampled})}{1 - P(D = 1|k, \text{ sampled})} \frac{n_{\bar{D}}}{n_D} \frac{1 - \rho}{\rho}.$$

The cutpoints are chosen such that v_k is the smallest number larger than or equal to $k/10$. Let $v_0 = 0$, the empirical predictiveness curve is generated as $R(v) = P(D = 1|k)$ for $v \in (v_{k-1}, v_k]$. Again, the predictiveness curves derived from parametric ROC models do not appear to be dramatically different from the empirical curve. The predictiveness curve derived from the whole bilomax ROC curve is very similar to that derived from the partial bilomax ROC curve whereas the predictiveness curve derived from the partial binormal ROC curve looks much more different. In general, the curve for PSA velocity is shallower, indicating that it is a poorer marker of risk for prostate cancer. Figure 4 shows the predictiveness curves for PSA and PSA velocity derived from the partial bilomax and binormal ROC curves with pointwise 95% percentile bootstrap confidence intervals. Comparing Figure 4(a) with Figure 4(b), we see that assuming different parametric ROC models has a big impact on both the predictiveness curve estimates and their variances, even if there does not appear to be a dramatic difference between the corresponding ROC curves (Figure 2(a) and (b)). This is consistent with our previous argument about the distinction between risk prediction and classification: one risk model can make a big difference compared to another in terms of risk prediction without having much impact in terms of classification.
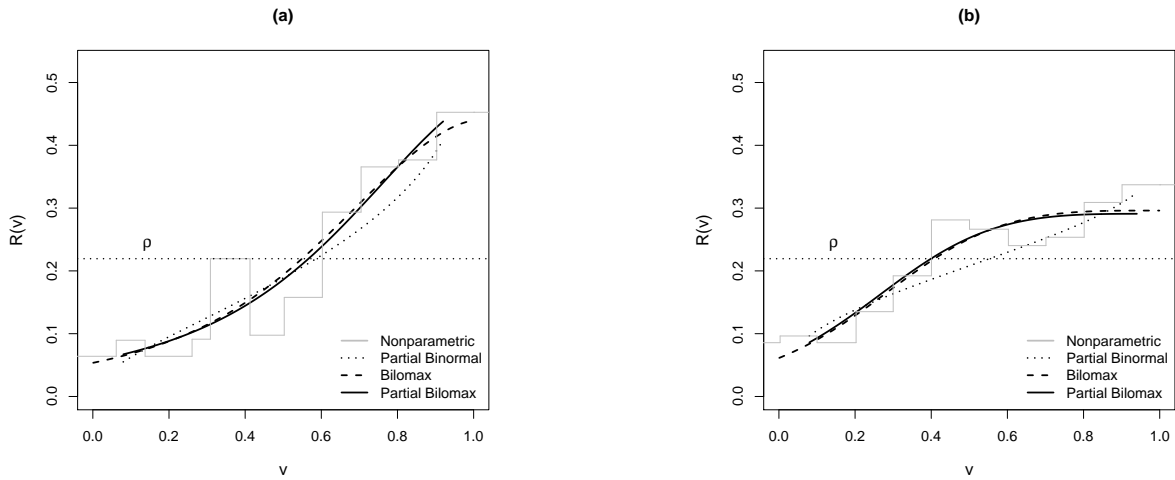
15

Figure 3. Predictiveness curves for two markers of prostate cancer estimated from the PCPT cohort and case-control data: (a) PSA, (b) PSA velocity.
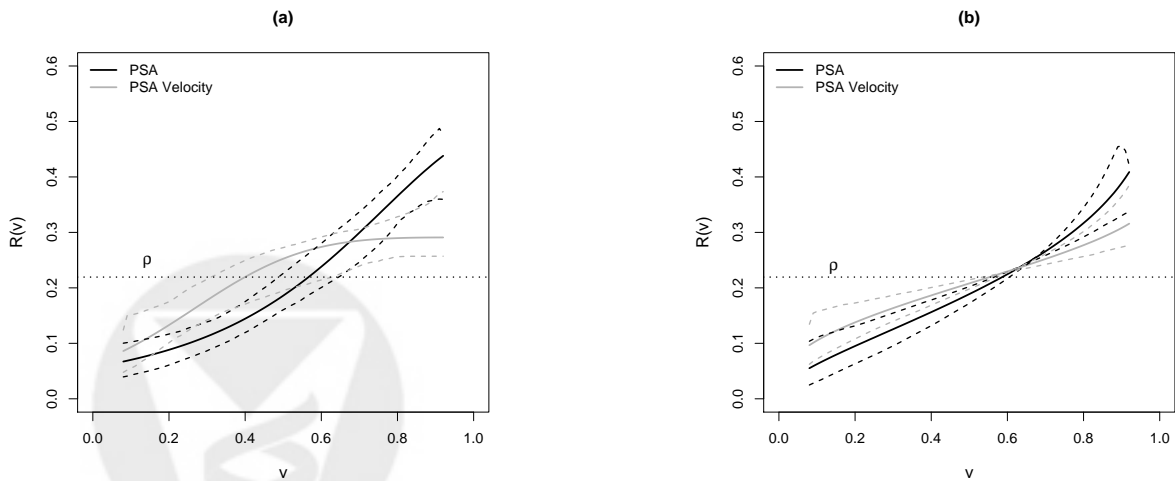

Figure 4. Predictiveness curves for two markers of prostate cancer estimated from the PCPT case-control data (solid lines) and their pointwise 95% percentile bootstrap confidence intervals (dashed lines): (a) bilomax, (b) partial binormal.

16

Table 2

Estimation of predictiveness curves for PSA and PSA velocity using case-control data from the PCPT study. Estimation are based on parametric ROC models. P-values are based on the comparison between PSA and PSA velocity using bootstrap variance estimates.

Model		PSA Est	PSA 95% CI	PSA Velocity Est	PSA Velocity 95% CI	p-value
Partial Bilomax	$R(0.1)$	7.0	(4.2, 10.2)	9.3	(5.5, 15.1)	0.378
	$R(0.9)$	42.7	(35.9, 48.0)	29.1	(25.7, 36.4)	< 0.001
	$R^{-1}(0.10)$	25.2	(11.5, 34.6)	11.9	(8.0, 21.1)	0.022
	$1 - R^{-1}(0.29)$	31.5	(23.2, 38.3)	18.6	(12.4, 42.2)	0.122
	$R^{-1}(0.29) - R^{-1}(0.10)$	43.3	(30.8, 61.8)	69.5	(36.0, 72.0)	0.020
Partial Binormal	$R(0.1)$	6.2	(3.1, 11.0)	10.5	(7.2, 15.7)	0.107
	$R(0.9)$	38.9	(33.0, 45.6)	30.8	(27.4, 36.5)	0.022
	$R^{-1}(0.10)$	21.6	(10.5, 31.3)	8.8	(8.1, 19.1)	0.018
	$1 - R^{-1}(0.29)$	24.9	(20.5, 26.9)	15.4	(8.2, 22.6)	0.010
	$R^{-1}(0.29) - R^{-1}(0.10)$	53.5	(42.6, 66.3)	75.8	(59.7, 81.1)	0.002

Table 2 presents estimates of $R(0.1)$ and $R(0.9)$. Also shown are proportions allocated into the low, high, and equivocal risk ranges, given a low risk threshold 11% and a high risk threshold 29%. For example, based on the partial bilomax model, the 90th percentile of risk is 0.291 according to PSA velocity while it is higher, 0.427, according to the absolute most recent PSA measurement ($p < 0.001$). At the low end of the scale, the 10th percentiles of risk based on PSA velocity and PSA are 0.07 and 0.093 respectively ($p = 0.378$). According to PSA velocity 14.5% of men can be classified as having risk below 11% while far more, 29.1%, qualify as low risk when using most recent PSA as the marker ($p = 0.009$). In addition, a greater fraction are found to have risks above 29% with PSA than with PSA velocity, 31.5% vs 18.6% ($p = 0.122$). Less men are categorized into the equivocal risk zone by PSA 39.4% than by PSA velocity 66.9% ($p = 0.014$). These results suggest that PSA performs better for stratifying risk of prostate cancer compared to PSA velocity. Inference based on the partial binormal ROC model is similar.

Finally, note that the parametric ROC curve methodology can be applied to a cohort study as well by plugging in the sample prevalence. Table 3 shows the results of predictiveness curves estimated using the entire PCPT cohort. We compare the bilomax ROC model based approach

17

Table 3

Estimation of predictiveness curves for PSA and PSA velocity using data for the entire PCPT cohort. The semi-parametric method of Huang et al. (2007) is compared with parametric ROC based method. P-values are based on the comparison between PSA and PSA velocity using bootstrap variance estimates.

Model		PSA		PSA Velocity		p-value
		Est	SE	Est	SE	
Semiparametric	$R(0.1)$	9.0	0.0067	14.5	0.0077	< 0.001
	$R(0.9)$	37.2	0.0092	29.5	0.0078	< 0.001
	$R^{-1}(0.10)$	13.6	0.020	3.0	0.0083	< 0.001
	$1 - R^{-1}(0.29)$	25.5	0.0099	11.1	0.018	< 0.001
	$R^{-1}(0.29) - R^{-1}(0.10)$	60.9	0.027	85.9	0.026	< 0.001
Bilomax	$R(0.1)$	9.2	0.0062	12.7	0.0068	< 0.001
	$R(0.9)$	39.5	0.011	33.8	0.010	< 0.001
	$R^{-1}(0.10)$	14.6	0.033	0	0.0059	< 0.001
	$1 - R^{-1}(0.29)$	28.4	0.017	22.3	0.018	0.007
	$R^{-1}(0.29) - R^{-1}(0.10)$	57.0	0.048	77.7	0.021	< 0.001

and the semi-parametric method developed in Huang et al. (2007) assuming $\text{logit}\{P(D = 1|Y)\} = \beta_0 + \beta_1 Y^{(\beta_2)}$ where $Y^{(\lambda)} = (Y^\lambda - 1)/\lambda$ when $\lambda \neq 0$ and $Y^{(\lambda)} = logY$ when $\lambda = 0$. The results are in close agreement. Moreover, standard errors of estimators based on the two methods are similar in magnitude. This suggests that by applying the case-control methods presented here to cohort data, efficiency is not lost relative to existing methods that only apply to cohort studies.

5. Discussion

Classification accuracy is usually considered to be an intrinsic property of a marker because it does not depend on the population-specific disease prevalence. Predictiveness, on the other hand, integrates classification accuracy and disease prevalence (Pepe et al., 2007) and characterizes the risk prediction ability of the marker in a particular population. In this paper, we show the one-to-one relationship between the ROC curve and the predictiveness curve when disease prevalence is known. The latter has been proposed as a graphical tool for evaluating a continuous risk prediction marker. We developed methodology for estimating the predictiveness curve based on a parametric ROC model from a case-control study design. The idea of estimating an ROC curve first seems

18

very natural in this retrospective setting considering that criteria for classification accuracy are defined conditional on disease status. The availability of a wide variety of methods for estimating a parametric ROC curve makes this approach even more appealing.

The main limitation of assuming a parametric ROC model, however, is lack of flexibility in the predictiveness curve estimator, especially at the boundary. Estimating a partial predictiveness curve from a portion of the ROC curve holds promise for resolving this issue. At the same time, it allows use of the most popular parametric ROC model, the binormal ROC model, which may not be concave in the whole range of FPF but can be restricted to be concave in certain regions.

Our methods can be used to compare the risk prediction capacities of different markers as we have shown with the prostate cancer example. Some extensions should be considered. For example, when predictiveness curves in subpopulations are of interest, we can estimate the covariate-specific ROC curve using existing ROC regression methods and derive the corresponding covariate-specific predictiveness curve by plugging in the disease prevalence in the subpopulation. Here we have focused on disease status at a fixed point in time. When subjects are observed over time, the time dimension may make things more challenging especially if there is censoring. How to incorporate the time dependence in a survival analysis setting requires further investigation.

So far we have assumed exact knowledge of the disease prevalence in the population or subpopulations. This is rarely true in real life. In practice, we obtain disease prevalence from the literature or estimate it from some pilot study. Sensitivity analyses that employ a range of plausible prevalence values should always be performed. In a more complicated two-phase design, disease prevalence estimated from the phase-one random sample could be used when we estimate the predictiveness curve using the phase-two case-control sample. Plugging in a prevalence estimate introduces extra variation in the predictiveness curve estimator which needs to accounted for in making inference. Further research in this regard is warranted.

19

Acknowledgments

References

Alonzo, T. A. and Pepe, M. S. (2002). Distribution-free ROC analysis using binary regression techniques. *Biostatistics* **3 (3),** 421-432.

Cai, T. and Moskowitz, C. S. (2004). Semi-parametric estimation of the binormal ROC curve for a continuous diagnostic test. *Biostatistics* **5(4),** 573-586.

Campbell, G. and Ratnaparkhi, M. V. (1993). An application of Lomax distributions in receiver operating characteristic (ROC) curve analysis. *Communications in Statistics* **22,** 1681-1697.

Cook, N. R. (2007). Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* **115,** 928-935.

Dodd, L. and Pepe, M. S. (2003). Semiparametric regression for the area under the receiver operating characteristic curve. *JASA* **98(462),** 409-417.

Dorfman, D. D. and Alf, E. (1969). Maximum likelihood estimation of parameters of signal detection theory and determination of confidence intervals-rating method data. *Journal of Mathematical Psychology* **6,** 487-496.

Dorfman, D. D. and Berbaum, K. S. and Metz, C. E. and Length, R. V. and Hanley, J. A. and Dagga, H. A. (1996). Proper receiver operating characteristic analysis: The bigamma model. *Academic Radiology* **4,** 138-149.

England, W. L. (1988). An exponential model used for optimal threshold selection in ROC curve. *Medical Decision Making* **8,** 120-131.

20

Gail, M. H. and Pfeiffer, R. M. (2005). On criteria for evaluating models of absolute risk. *Biostatistics* **6(2),** 227-239.

Greenhouse, S. W. and Mantel, N. (1950). The evaluation of diagnostic tests. *Biometrics* **6,** 399-412.

Hanley, J. A. (1988). The robustness of the 'binormal' assumptions used in fitting ROC curves. *Medical Decision Making* **143,** 29-36.

Hsieh, F. and Turnbull, B. W. (1996). Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *Annals of Statistics* **24,** 25-40.

Huang, Y. and Pepe, M. S. and Feng, Z. (2007). Evaluating the predictiveness of a continuous marker. *Biometrics* (In Press).

Jiang, Y. and Metz, C. E. and Nishikawa, R. M. (1996). A receiver operating characteristic partial area index for highly sensitive diagnostic tests. *Radiology* **201,** 745-750.

Lomax, H. S. (1954). Business failure; Another example of the analysis of failure data. *JASA* **49,** 847-852.

McClish, D. K. (1989). Analyzing a portion of the ROC curve. *Medical Decision Making* **9,** 190-195.

McIntosh, M. and Pepe, M. S. (2002). Combining several screening tests: Optimality of the risk score. *Biometrics* **58,** 657-664.

Metz, C. and Herman, B. A. and Shen, J. H. (1998). Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously distributed data. *Statistics in Medicine* **17,** 1033-1053.

Metz, C. E. and Pan, X. (1999) "Proper" binormal ROC curves: Theory and maximum likelihood estimation. *Journal of Mathematical Psychology 43,* 1-33.

Obuchowski, N. A. (2003). Receiver operating characteristic curves and their use in radiology. *Radiology* **229,** 3-8.

Pepe, M. S. (2000). An interpretation for the ROC curve and inference using GLM procedures. *Biometrics* **56,** 352-359.

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction.* Oxford University Press.

Pepe, M. S. and Cai, T. (2004). The analysis of placement values for evaluating discriminatory measures. *Biometrics* **60,** 528-535.

Pepe, M. S. and Etzioni, R. and Feng, Z. and Potter, J. D. and Thompson, M. L. and Thornquist, M. and Winget, M. and Yasui, Y. (2001) Phases of biomarker development for early detection of cancer. *Journal of the National Cancer Institute* **93(14),** 1054-1061.

Pepe, M. S. and Feng, Z. and Huang, Y. and Longton, G. M. and Prentice, R. and Thompson, I. M. and Zheng, Y. (2007). Integrating the predictiveness of a marker with its performance as a classifier. *American Journal of Epidemiology* (In Press).

Thompson, M. L. and Zucchini, W. (1989). On the statistical analysis of ROC curves. *Statistics in Medicine* **8,** 1277-1290.

Thompson, I. M. and Pauler Ankerst, D. and Chi, C. (2006). Assessing prostate cancer risk: results from the prostate cancer prevention trial. *Journal of the National Cancer Institute* **98,** 529-534.

Tosteson, A. N. A. and Begg, C. B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making* **3,** 204-215.

Zou, K. H. and Hall, W. J. (2000). Two transformation models for estimating an ROC curve derived from continuous data. *Journal of Applied Statistics* **27,** 621-631.

Wieand, S. and Gail, M. H. and James, B. R. and James, K. L. (1989). A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika* **76,** 585-592.

22