Adjusting for covariate effects on classification accuracy using the covariate-adjusted receiver operating characteristic curve

BY HOLLY JANES AND MARGARET S. PEPE

*Division of Public Health Sciences, Fred Hutchinson Cancer
Research Center, 1100 Fairview Avenue North,
Seattle, Washington 98109, U.S.A.*

hjanes@scharp.org mspepe@u.washington.edu

SUMMARY

Recent scientific and technological innovations have produced an abundance of potential markers that are being investigated for their use in disease screening and diagnosis. In evaluating these markers, it is often necessary to account for covariates associated with the marker of interest. Covariates may include subject characteristics, expertise of the test operator, test procedures or aspects of specimen handling. In this paper, we propose the covariate-adjusted receiver operating characteristic curve, a measure of covariate-adjusted classification accuracy. Nonparametric and semiparametric estimators are proposed, asymptotic distribution theory is provided and finite sample performance is investigated. For illustration we characterize the age-adjusted discriminatory accuracy of prostate-specific antigen as a biomarker for prostate cancer.

Some key words: Classification accuracy; Covariate effect; Receiver operating characteristic curve; Sensitivity; Specificity.

1. INTRODUCTION

Research into new markers for disease diagnosis, screening and prognosis has exploded in recent years. The primary question in each setting is of classification accuracy: how well does the marker distinguish between the two groups of individuals, the ‘cases’ and the ‘controls’?

The receiver operating characteristic or ROC curve plays a central role in evaluating classification accuracy (Baker, 2003; Pepe et al., 2001). Let D denote the binary group variable such as disease status, and let Y_D and $Y_{\bar{D}}$ denote case and control marker observations with survivor functions $S_D(y) = \text{pr}(Y_D > y)$ and $S_{\bar{D}} = \text{pr}(Y_{\bar{D}} > y)$. The ROC curve is a plot of the true-positive fraction or sensitivity versus the false-positive fraction or 1-specificity for rules that classify an individual as ‘test-positive’ if $Y > c$, where the threshold c varies over all possible values. Equivalently, if the false-positive fraction is t , then the receiver operating characteristic curve is $\text{ROC}(t) = \text{pr}\{Y_D > S_{\bar{D}}^{-1}(t)\} = S_D\{S_{\bar{D}}^{-1}(t)\}$ (Pepe, 2003, p. 69).

There are often factors, such as patient or disease characteristics or features of the specimen handling or test procedure, which affect the marker and its associated classification accuracy. These effects may mean that the definition of testing positive on the basis of the marker should depend on covariates, or that the accuracy of the test is less than optimal in certain settings (Pepe, 2003, pp. 48–9). This paper proposes a covariate-adjusted summary of classification accuracy, the covariate-adjusted ROC curve.

2. THE COVARIATE-ADJUSTED ROC CURVE

Let Y be a continuous marker and let Z be a covariate. The methods generalize naturally to multiple covariates. Let Z_D and $Z_{\bar{D}}$ denote case and control covariate observations with cumulative distribution functions P_D and $P_{\bar{D}}$. Denote by $S_D(y | z) = \text{pr}(Y_D > y | Z_D = z)$ and $S_{\bar{D}}(y | z) = \text{pr}(Y_{\bar{D}} > y | Z_{\bar{D}} = z)$ the continuous survivor functions for Y conditional on $Z = z$, where $S_D^{-1}(t | z)$ and $S_{\bar{D}}^{-1}(t | z)$ are the inverse conditional survivor functions, and let $f_D(y | z)$ and $f_{\bar{D}}(y | z)$ be the conditional densities.

The covariate-adjusted ROC curve at a false-positive fraction of t is the overall true-positive fraction when thresholds for defining ‘test-positive’ are covariate-specific, chosen to ensure that the value of the false-positive fraction is t in each covariate-specific population. Mathematically, we write $\text{AROC}(t) = \text{pr}\{Y_D > S_{\bar{D}}^{-1}(t | Z_D)\}$, where we note that the threshold equal to the quantile $S_{\bar{D}}^{-1}(t | Z_D)$ yields a false-positive fraction of t in the covariate-specific population. Naturally, the marginal false-positive fraction is also equal to t .

The covariate-specific receiver operating characteristic curve, $\text{ROC}(t | z) = S_D\{S_{\bar{D}}^{-1}(t | z) | z\}$, is the ROC curve for the marker in the population with covariate level $Z = z$. Interestingly, the adjusted ROC can be written as a weighted average of covariate-specific curves,

$$\begin{aligned} \text{AROC}(t) &= \int \text{pr}\{Y_D > S_{\bar{D}}^{-1}(t | Z_D) | Z_D = z\} dP_D(z) \\ &= \int \text{ROC}(t | z) dP_D(z). \end{aligned} \quad (1)$$

Equivalently, $\text{AROC}(t) = E\{\text{ROC}(t | Z_D)\}$, where the expectation is taken with respect to Z_D . Equation (1) shows that when Z affects marker observations but not discriminatory accuracy as defined by the ROC curve, the adjusted ROC is the common covariate-specific ROC curve; it is perhaps best motivated in this setting. On the other hand, when Z affects discrimination, the adjusted ROC reports a weighted average of covariate-specific true-positive fractions, holding the covariate-specific false-positive fractions constant. While estimation of covariate-specific ROC curves is typically of primary interest in this context, the adjusted ROC curve is a summary of covariate-adjusted accuracy that is useful for comparing markers and in small studies where covariate-specific ROC curves cannot be estimated with precision.

Another interpretation follows from noting that $\text{AROC}(t) = \text{pr}\{S_{\bar{D}}(Y_D | Z_D) \leq t\}$, the cumulative distribution function of $S_{\bar{D}}(Y_D | Z_D)$, where $S_{\bar{D}}(Y_D | Z_D)$ is the placement of a case observation relative to a reference distribution for controls with the same covariate value. Contrast this with the unadjusted or pooled ROC curve which has been shown to be equal to the cumulative distribution function of a case observation standardized relative to the general control distribution (Pepe & Cai, 2004), $\text{ROC}(t) = \text{pr}\{S_{\bar{D}}(Y_D) \leq t\}$.

Expression (1) has some attractive mathematical properties, including invariance with respect to monotone increasing transformations of Y and/or Z . It is also unaffected by control covariate-dependent sampling. This follows because such designs sample controls randomly conditional on Z , and cases randomly marginally. Janes & Pepe (2008) argue for its use in studies that employ frequency matching, where controls are sampled to have the same distribution of Z as cases. This is one type of covariate-dependent sampling.

3. ESTIMATION

3.1. Estimators

We propose two estimators for $\text{AROC}(t) = \text{pr}\{Y_D > S_{\bar{D}}^{-1}(t | Z_D)\}$ using n_D and $n_{\bar{D}}$ case and control observations. In both instances, the outside probability is estimated empirically. With the

nonparametric estimator, valid for a discrete covariate $Z \in \{1, \dots, K\}$, we estimate the quantiles $S_{\bar{D}}^{-1}(t | z)$ empirically in each covariate stratum. With the semiparametric estimator, the quantiles are estimated based on a model for $Y_{\bar{D}}$ as a function of $Z_{\bar{D}}$. We lay out the general framework for the adjusted ROC estimator, of which the nonparametric estimator is a special case.

We assume the quantile model $Y_{\bar{D}} = f(Z_{\bar{D}}, \epsilon; \theta)$, where ϵ is random error and θ are parameters. With the semiparametric adjusted ROC estimator, this model may be parametric, such as a normal linear model, or semiparametric (Heagerty & Pepe, 1999). Let $q(t; z, \theta) = S_{\bar{D}}^{-1}(t | z; \theta)$ be the function which extracts the $1 - t$ quantile from the set of control quantiles with covariate value $Z = z$. We write

$$\text{AROC}_{\hat{\theta}}(t) = n_D^{-1} \sum_{i=1}^{n_D} I\{Y_{Di} > q(t; Z_{Di}, \hat{\theta})\}.$$

With the nonparametric estimator, $\theta = \{S_{\bar{D}}^{-1}(t | z = 1), \dots, S_{\bar{D}}^{-1}(t | z = K)\}^T$ are the quantiles themselves, $\hat{S}_{\bar{D}}(y | z) = n_{\bar{D}}(z)^{-1} \sum_{i=1}^{n_{\bar{D}}(z)} I\{Y_{\bar{D}_i} > y, Z_{\bar{D}_i} = z\}$, and $q(t; z, \hat{\theta}) = \hat{S}_{\bar{D}}^{-1}(t | z) = \inf_{s \in [0, 1]} \{\hat{S}_{\bar{D}}(s | z) \geq t\}$, all estimated using $n_{\bar{D}}(z)$ observations in each covariate stratum. This estimator depends only on the ranks of the data, and thus is invariant with respect to monotone transformations.

3.2. Asymptotic distribution theory

We make the following assumptions in establishing asymptotic distribution theory. Recall that the distribution of Y_D is not a function of θ .

Assumption 1. We assume that observations are randomly sampled conditional on D , that $n_D + n_{\bar{D}} \rightarrow \infty$ and that $n_D/n_{\bar{D}} \rightarrow \lambda \in (0, 1)$.

Assumption 2. We assume that $n_D^{1/2}(\hat{\theta} - \theta)$ converges in distribution to a normal zero-mean random variable with covariance matrix Σ_{θ} as $n_{\bar{D}} \rightarrow \infty$.

Assumption 3. We assume that $\text{AROC}_{\theta}(t)$ is differentiable, and hence continuous, in θ .

Assumption 4. We assume that $\lim_{n_{\bar{D}} \rightarrow \infty} \text{pr}\{\text{AROC}_{\hat{\theta}}(t) \notin \{0, 1\}\} = 1$, where $\text{AROC}_{\hat{\theta}}(t) = \text{pr}\{Y_D > q(t; Z_D, \hat{\theta}) | \hat{\theta}\}$ is the adjusted ROC based on estimated quantiles.

Assumption 5. We assume that $t \notin \{0, 1\}$.

In relation to Assumption 1, covariate-dependent sampling can also be accommodated; see § 3.4. A wide variety of quantile models satisfy Assumption 2, including parametric models (Cole, 1990; Cole & Green, 1992; Pepe, 2003, p. 140), semiparametric models (Heagerty & Pepe, 1999; Zheng & Heagerty, 2004), empirical methods, as proven in the Appendix, and any $\hat{\theta}$ based on unbiased estimating equations satisfying standard regularity conditions. Assumption 3 is also valid for a diversity of quantile and ROC models, such as the location-scale quantile model (Heagerty & Pepe, 1999) with bounded $(\partial/\partial t) \text{ROC}(t | z)$ and $E(Z_D) < \infty$; see an unpublished University of Washington technical report by H. Janes and M. S. Pepe, available at <http://www.bepress.com/uwbiostat/paper283>. Assumption 4 is violated if the support for Y_D is entirely above or below the estimated quantile of interest. This will not occur as long as the support for Y_D includes the support for $Y_{\bar{D}}$ or if the support for Y_D is unbounded, as under the normal distribution. We also require that $t \notin \{0, 1\}$, but by definition $\text{AROC}(0) = 0$ and $\text{AROC}(1) = 1$.

Finally, imposing continuity of $S_D(y | z)$ and $S_{\bar{D}}(y | z)$ implies that $\text{ROC}(t | z)$ and $\text{AROC}(t)$ are continuous in t . The proof of the following theorem is given in the Appendix.

THEOREM 1. *Under Assumptions 1–5, $n_D^{1/2} \{\text{AROC}_{\hat{\theta}}(t) - \text{AROC}_{\theta}(t)\}$ converges in distribution to a normal zero-mean random variable with variance $V(t)$ as $n_D, n_{\bar{D}} \rightarrow \infty$, where*

$$V(t) = \text{AROC}_{\theta}(t) \{1 - \text{AROC}_{\theta}(t)\} + \lambda \frac{\partial}{\partial \theta} \text{AROC}_{\theta}(t) \Sigma_{\theta} \frac{\partial}{\partial \theta^T} \text{AROC}_{\theta}(t). \quad (2)$$

The form of $V(t)$ is intuitively plausible. The second component comes from estimating the Z -specific quantiles, while the first is a binomial variance associated with estimating the true-positive fraction, given the quantiles.

For the nonparametric estimator, Assumptions 2 and 3 are satisfied when the following assumption holds.

Assumption 6. We assume that $f_{\bar{D}}(y | z)$ is continuous and positive in a neighbourhood of $S_{\bar{D}}^{-1}(t | z)$ for all z .

Under Assumptions 1–6, $V(t)$ reduces to

$$V(t) = \text{AROC}_{\theta}(t) \{1 - \text{AROC}_{\theta}(t)\} + \lambda \left[\sum_{z=1}^K \frac{p_D^2(z)}{p_{\bar{D}}(z)} \frac{f_D\{S_{\bar{D}}^{-1}(t | z) | z\}^2}{f_{\bar{D}}\{S_{\bar{D}}^{-1}(t | z) | z\}^2} \right] t(1 - t), \quad (3)$$

where p_D and $p_{\bar{D}}$ are the probability mass functions for Z_D and $Z_{\bar{D}}$; see the Appendix for a proof.

3.3. Consistent variance estimation

Our first variance estimator can be used to estimate the variance of the semiparametric adjusted ROC estimator, namely expression (2) multiplied by n_D^{-1} . The semiparametric estimator is consistent by Theorem 1. We assume that a consistent estimator of Σ_{θ} exists; for example, if $\hat{\theta}$ is based on a set of unbiased estimating equations, a sandwich-type variance estimator can be used. The j th component of $(\partial/\partial \theta) \text{AROC}_{\theta}(t)$ is estimated by $(\text{AROC}_{\hat{\theta}+h_j(n)} - \text{AROC}_{\hat{\theta}-h_j(n)})/\{2h(n)\}$, where $h(n)$ is $o(n_D^{-1/3})$, and $\hat{\theta} + h_j(n)$, respectively $\hat{\theta} - h_j(n)$, denotes the vector $\hat{\theta}$ with $h(n)$ added to, respectively subtracted from, the j th component only. In the Appendix, the composite variance estimator is shown to be consistent under Assumptions 1–5 and the following Assumption 7.

Assumption 7. We assume that

$$\lim_{n_D \rightarrow \infty} \text{pr}\{\text{AROC}_{\hat{\theta}+h_j(n)} \notin \{0, 1\}\} = \lim_{n_{\bar{D}} \rightarrow \infty} \text{pr}\{\text{AROC}_{\hat{\theta}-h_j(n)} \notin \{0, 1\}\} = 1,$$

for all j .

With small sample sizes, the estimate of $(\partial/\partial \theta_j) \text{AROC}_{\theta}(t)$ may be sensitive to the choice of bandwidth, $h(n)$. We have used $h(n) = 0.04$ in applications and simulations; this value ensured that $(\text{AROC}_{\hat{\theta}+h_j(n)} - \text{AROC}_{\hat{\theta}-h_j(n)})/\{2h(n)\} \simeq (\partial/\partial \theta_j) \text{AROC}_{\theta}(t)$ in one example and has worked well in practice. We leave exploration of the optimal choice of $h(n)$ for future research.

Our second variance estimator can be used to estimate the variance of the nonparametric estimator, namely expression (3) multiplied by n_D^{-1} . Here, $\text{AROC}_{\theta}(t)$ is estimated using the nonparametric estimator, p_D and $p_{\bar{D}}$ by binomial proportions, $S_{\bar{D}}^{-1}(t | z)$ empirically, and $f_D(y | z)$ and $f_{\bar{D}}(y | z)$ with uniformly consistent kernel density estimators (Silverman, 1986, § 3.7). In the Appendix we prove that the composite function is consistent under Assumptions 1–6 and the following Assumption 8.

Assumption 8. We assume that $f_D(y | z)$ and $f_{\bar{D}}(y | z)$ are continuous density functions for all z .

Bootstrap variance estimation is a simple alternative which accommodates clustered sampling as well as performing well in practice and in simulations; see § 4.

3.4. Sampling based on covariates

In many situations, sampling may depend on both D and Z . Two examples of this are frequency matching and sampling subjects in a specified range of Z . Our asymptotic results hold under such designs, but all population distributions should be replaced with sampling distributions.

3.5. Estimation using ROC regression

The AROC can also be estimated using the ROC regression method of Pepe (2000) and Alonzo & Pepe (2002), which requires estimation of covariate-specific control quantiles and specifying and fitting a model for the ROC curve, typically as a function of covariates. A model for the adjusted ROC curve is obtained by including Z in the quantile calculations, while omitting Z from the ROC model. A classic example is the binormal model, $\text{AROC}(t) = \Phi\{\alpha + \beta\Phi^{-1}(t)\}$, where Φ is the standard normal cumulative distribution function. This approach results in a smooth parametric estimate of the adjusted ROC, but the marker distributions remain unspecified. The semiparametric ROC regression estimator of Cai & Pepe (2002) reduces to our semiparametric estimator.

4. SMALL SAMPLE PERFORMANCE

We evaluate the finite sample properties of our estimators using simulations, starting with the nonparametric estimator and its variance which can be used for discrete Z . We assume that there exists a monotone increasing transformation under which Y is normally distributed conditional on a binary covariate, Z , in both cases and controls: $Y_{\bar{D}} \sim N(0, 1)$ and $Y_D \sim N(0.9, 1)$ conditional on $Z = 0$; $Y_{\bar{D}} \sim N(0.2, 1)$ and $Y_D \sim N(0.9, 1)$ conditional on $Z = 1$; $\text{pr}(Z_{\bar{D}} = 1) = 0.7$ and $\text{pr}(Z_D = 1) = 0.3$. All of the assumptions laid out in § 3 are satisfied under this model. The values of $\text{AROC}(t)$ at $t = 0.05, 0.10, 0.20$ and 0.50 are $0.21, 0.33, 0.50$ and 0.80 and the two components of asymptotic variance are $n_D^{-1}(0.33, 0.44, 0.50, 0.32)$ and $n_{\bar{D}}^{-1}(1.41, 1.40, 1.14, 0.40)$.

We simulated 5000 datasets, where $n_D = n_{\bar{D}}$ varies between 100 and 1000; see Table 1. In terms of percentage bias, defined as $100[\text{avg}\{\widehat{\text{AROC}}_{\hat{\theta}}(t)\} - \text{AROC}_{\theta}(t)]/\text{AROC}_{\theta}(t)$, where $\text{avg}\{\widehat{\text{AROC}}_{\hat{\theta}}(t)\}$ is the average estimate, the estimator performs very well, except for some modest bias when both t and $n_D = n_{\bar{D}}$ are small. The percentage bias in the nonparametric variance estimator, based on rectangular kernel density estimates, is $100[n_D^{-1}\text{med}\{\hat{V}(t)\} - \text{var}\{\widehat{\text{AROC}}_{\hat{\theta}}(t)\}]/\text{var}\{\widehat{\text{AROC}}_{\hat{\theta}}(t)\}$, where the median is calculated because of the skewed distribution of the variance estimates, and $\text{var}\{\widehat{\text{AROC}}_{\hat{\theta}}(t)\}$ is the sample variance of the estimates. The variance estimator tends to underestimate the true variance. There is substantial bias when t is small, mostly due to estimation of the second component of variance, but this disappears for larger t . The percentage difference between the asymptotic and sample variances, $100[n_D^{-1}V(t) - \text{var}\{\widehat{\text{AROC}}_{\hat{\theta}}(t)\}]/\text{var}\{\widehat{\text{AROC}}_{\hat{\theta}}(t)\}$, tends to be small, with differences only when both t and $n_D = n_{\bar{D}}$ are small. Coverage probabilities using nonparametric variance estimators are provided. Coverage based on logit transformations, which have been shown to improve coverage for the pooled ROC when t is close to 0 or 1 (Pepe, 2003, p. 102), are also shown. Only logit-based coverage is shown when both t and $n_D = n_{\bar{D}}$ are small, as adjusted ROC estimates are close to zero. We find that coverage can be low for small t but is very good for moderate t .

Table 1. *Small sample performance of the nonparametric estimator of the adjusted ROC based on 5000 simulations. The sample size, $n_D = n_{\bar{D}}$, varies between 100 and 1000. The nonparametric variance estimator uses rectangular kernel density estimates. Results are shown at four false-positive fractions of interest. Only coverage based on logit-transformations is shown for small t and $n_D = n_{\bar{D}}$, since $\text{AR}\hat{\text{C}}_{\hat{\theta}}(t)$ is frequently close to zero. Bootstrap variance estimates based on 100 bootstrap samples and associated coverage are also shown*

n_D	% Bias	% Diff	% Bias var	Cov	Logit cov	% Bias Boot var	Cov	Logit cov
$t = 0.05$								
100	6.15	21.34	-30.65	—	89.36	-25.25	—	87.32
200	10.05	7.83	-29.67	—	88.14	-6.30	—	91.68
500	2.40	6.88	-20.25	—	91.58	4.36	—	94.86
1000	1.71	2.28	-16.24	91.90	92.14	2.92	94.40	94.18
$t = 0.10$								
100	7.83	15.48	-19.75	—	91.22	-7.32	—	93.38
200	4.37	7.06	-17.53	—	91.88	0.79	—	94.82
500	0.75	3.88	-12.94	—	93.22	2.27	—	94.60
1000	0.94	1.39	-10.53	92.86	92.18	1.68	94.10	94.40
$t = 0.20$								
100	2.21	12.70	-4.42	91.92	95.36	1.04	93.88	97.08
200	1.45	6.64	-6.00	92.76	94.38	3.08	94.26	95.68
500	0.24	3.51	-3.96	93.74	94.42	2.96	94.44	95.08
1000	0.33	4.26	-1.12	94.14	94.42	5.04	94.16	94.56
$t = 0.50$								
100	0.02	2.05	-3.67	92.16	96.98	8.41	95.00	97.26
200	-0.11	-0.10	-0.47	93.42	95.86	4.38	94.18	95.98
500	-0.05	-1.23	1.78	94.78	95.98	0.71	94.64	95.42
1000	-0.01	2.48	6.37	95.36	95.88	3.76	94.78	95.20

% Bias, percentage bias in $\text{AROC}_{\theta}(t)$; % Diff, percentage difference between asymptotic and sample variances; % Bias var, percentage bias in the nonparametric variance estimator; Cov, coverage of 95% confidence intervals; Logit cov, coverage based on logit-transformations; Boot var, bootstrap variance estimates.

We also evaluate the performance of bootstrap variance estimates. Data are resampled 100 times conditional on D , and the sample variance of the adjusted ROC estimates is calculated. The bootstrap variance estimator exhibits substantially less bias than the nonparametric estimator. Bootstrap coverage also tends to be better; coverage is good except when both t and $n_D = n_{\bar{D}}$ are small.

Table 2 compares the nonparametric estimator of the adjusted ROC with the semiparametric estimator based on a normal linear quantile model. The percentage difference in the estimates, defined as $\text{avg}\{\text{AR}\hat{\text{C}}_{\hat{\theta}; \text{semi}}(t) - \text{AR}\hat{\text{C}}_{\hat{\theta}}(t)\} / \text{AROC}_{\theta}(t)$, where $\text{AR}\hat{\text{C}}_{\hat{\theta}}(t)$ is the nonparametric estimator and $\text{AR}\hat{\text{C}}_{\hat{\theta}; \text{semi}}(t)$ is the semiparametric estimator, tends to be small. The estimated relative efficiency of the two estimators, $\text{var}\{\text{AR}\hat{\text{C}}_{\hat{\theta}; \text{semi}}(t)\} / \text{var}\{\text{AR}\hat{\text{C}}_{\hat{\theta}}(t)\}$, where the variances are estimated from the 5000 simulations, shows that the semiparametric estimator yields substantial gains in efficiency, with larger gains for smaller t and larger $n_D = n_{\bar{D}}$.

The performances of the semiparametric estimator and its variance are explored under the double binormal model; see Lin & Jeon (2003) and the authors' technical report. Under this model, there is a monotone increasing transformation under which (Y, Z) is bivariate normal in both cases and controls. We set the parameter values to $E(Y_{\bar{D}}) = E(Z_{\bar{D}}) = 0$, $E(Y_D) = 0.7$, $E(Z_D) = 0.5$, $\text{var}(Y_{\bar{D}}) = \text{var}(Y_D) = 1$, $\text{var}(Z_{\bar{D}}) = \text{var}(Z_D) = 1.5^2$, $\text{corr}(Y_D, Z_D) = 0.6$ and $\text{corr}(Y_{\bar{D}}, Z_{\bar{D}}) = 0.2$. This is an extension of the classic binormal model for the pooled ROC curve (Swets, 1986;

Table 2. Comparison of the nonparametric and semiparametric estimates of the adjusted ROC based on 5000 simulations. The sample size, $n_D = n_{\bar{D}}$, varies between 100 and 1000. The semiparametric estimator uses a normal linear quantile model. The percentage difference in the adjusted ROC estimates and relative efficiency are shown at four false-positive fractions of interest

n_D	$t = 0.05$		$t = 0.10$		$t = 0.20$		$t = 0.50$	
	% Diff	RE	% Diff	RE	% Diff	RE	% Diff	RE
100	-0.20	0.61	-5.33	0.68	-2.24	0.76	-0.75	0.82
200	-6.87	0.54	-2.98	0.64	-1.25	0.72	-0.33	0.81
500	-1.90	0.51	-0.88	0.61	-0.47	0.71	-0.17	0.79
1000	-0.92	0.50	-0.52	0.61	-0.12	0.71	-0.04	0.82

% Diff, percentage difference in the adjusted ROC estimates; RE, relative efficiency.

Hanley, 1988, 1996). The induced adjusted ROC is a binormal ROC curve, as shown in the authors' technical report. All of the assumptions laid out in § 3 are satisfied under this model. We apply the semiparametric estimator using a normal linear quantile model; this is the true model for $Y_{\bar{D}}$ given $Z_{\bar{D}}$. The values of $\text{AROC}(t)$ at $t = 0.05, 0.10, 0.20$ and 0.50 are $0.16, 0.25, 0.39$ and 0.67 . The two components of asymptotic variance are $n_D^{-1}(0.24, 0.37, 0.49, 0.36)$ and $n_{\bar{D}}^{-1}(0.35, 0.53, 0.56, 0.23)$.

We simulated 5000 datasets, in which $n_D = n_{\bar{D}}$ varies between 100 and 1000. The estimator performs very well, except for some modest bias for very small $n_D = n_{\bar{D}}$ and t ; see Table 3. The semiparametric variance estimator exhibits moderate small sample bias for the smallest sample sizes; the variance is consistently overestimated. This is primarily due to bias in the second component of variance, which involves $(\partial/\partial\theta) \text{AROC}_{\theta}(t)$. However, coverage is reasonable. The asymptotic and sample variances agree quite well, except for some minor differences with the smallest sample sizes. Bootstrap variance estimates are good alternatives: they tend to exhibit less bias and have excellent coverage.

5. ILLUSTRATION

Data from the Physicians' Health Study (Gann et al., 1995) are used for illustration. This was a randomized controlled study of aspirin and β -carotene among 22 071 U.S. male physicians of ages 40 to 84 years in 1982. A blood sample taken at enrolment was stored. For 429 men diagnosed with prostate cancer up to 12 years after enrolment, most before PSA, Prostate-Specific Antigen, was widely used for screening, and for 1287 controls not diagnosed with prostate cancer during 12 years of follow-up, the serum was assayed for PSA. Cases and controls were matched on age; for each case, three controls within one year of age were selected (Gann et al., 1995; Etzioni et al., 2004).

The goal of this sub-study is to assess the ability of PSA to discriminate between men who did and did not develop prostate cancer. The pooled ROC curve in the matched data is not of practical interest (Janes & Pepe, 2008). It describes the ability of PSA to distinguish between cases and age-matched controls, an artificial control group. More importantly, this curve is attenuated by the matching on age. We use the adjusted ROC to quantify the age-adjusted classification accuracy of PSA.

Age-specific ROC curves for PSA, estimated using a binormal ROC regression model (Alonzo & Pepe, 2002) with quantiles from a linear location-scale model (Heagerty & Pepe,

Table 3. *Small sample performance of the semiparametric estimator of the adjusted ROC based on 5000 simulations. The sample size, $n_D = n_{\bar{D}}$, varies between 100 and 1000. The quantiles are estimated using a normal linear model, and the semiparametric variance estimator uses a bandwidth of $h = 0.04$. Results are shown at four false-positive fractions of interest. Only coverage based on logit transformations is shown for small t and $n_D = n_{\bar{D}}$, since $\text{AR}\hat{\text{OC}}_{\hat{\theta}}(t)$ is frequently close to zero. Bootstrap variance estimates based on 100 bootstrap samples and associated coverage are also shown*

n_D	% Bias	% Diff	% Bias var	Cov	Logit cov	% Bias Boot var	Cov	Logit cov
$t = 0.05$								
100	14.35	-7.64	1.50	-	91.52	6.25	-	93.20
200	6.47	0.28	5.17	-	93.88	6.67	-	94.64
500	2.97	0.21	3.70	-	94.60	1.97	-	95.40
1000	0.85	1.64	1.95	94.32	95.08	2.20	94.42	95.30
$t = 0.10$								
100	7.08	2.27	8.25	-	93.30	2.47	-	94.96
200	3.23	3.26	4.87	-	94.28	3.79	-	95.48
500	1.34	-2.22	-2.72	-	94.22	-3.09	-	94.82
1000	0.34	2.47	1.48	94.66	95.10	2.48	95.00	95.44
$t = 0.20$								
100	2.46	5.14	9.02	92.54	94.02	-0.12	93.50	95.86
200	1.26	2.31	4.54	93.80	94.70	0.25	94.38	95.54
500	0.37	-0.55	-0.26	94.10	95.42	-1.71	94.30	94.72
1000	0.03	-0.33	0.79	94.48	94.72	-0.17	94.42	94.66
$t = 0.50$								
100	-0.62	3.66	13.89	93.52	95.98	9.06	94.70	96.76
200	-0.44	1.08	9.75	94.58	95.44	4.63	94.92	95.68
500	-0.21	1.38	5.50	95.02	95.26	4.35	94.82	95.18
1000	-0.14	0.45	3.13	95.00	95.24	3.24	94.68	94.88

% Bias, percentage bias in $\text{AROC}_{\hat{\theta}}(t)$; % Diff, percentage difference between asymptotic and sample variances; % Bias var, percentage bias in the nonparametric variance estimator; Cov, coverage of 95% confidence intervals; Logit cov, coverage based on logit-transformations; Boot var, bootstrap variance estimates.

1999), reveal very little variation in discrimination due to age; see Fig. 1(a). Hence, the adjusted ROC represents the common age-specific ROC curve for PSA.

The adjusted ROC for PSA is shown in Fig. 1(b), estimated using both the semiparametric estimator and a binormal ROC regression model, where the control quantiles are estimated using a linear location-scale model (Heagerty & Pepe, 1999) for both methods. Bootstrapping is used for inference, and logit-based confidence intervals for the semiparametric estimator are overlaid at $t = 0.025$ and $t = 0.05$. The curve describes the ability of PSA to discriminate between cases and controls of the same age. We estimate that when the age-specific false-positive fraction is held at 0.025, 17% of cases can be detected, with a 95% confidence interval of 13% to 21%. When the common false-positive fraction is increased to 0.05, 27% of cases can be detected, with a 95% confidence interval of 21% to 33%.

6. DISCUSSION

When covariates affect discrimination, there are various ways of combining covariate-specific ROC curves. The adjusted ROC curve is a simple vertical average. A horizontal average may be more appropriate in certain settings, and is a simple extension of our methods.

Fig. 1. ROC curves for PSA in the Physicians' Health Study data. (a) Age-specific ROC curves, 50 (dotted), 60 (dash-dotted), 70 (dashed), estimated using the ROC regression model. (b) The age-adjusted ROC curve, estimated using the semiparametric estimator (the dotted line) and the ROC regression model (the solid line). The 95% confidence intervals, based on bootstrapped variance estimates, are overlaid at $t = 0.025$ and $t = 0.05$.

The area under the adjusted ROC, equivalent to $\text{pr}(Y_D > Y_{\bar{D}} \mid Z_D = Z_{\bar{D}})$, can be interpreted as the probability of correctly ordering a randomly chosen case and a control observation with the same covariate value. This statistical summary deserves further development and may be used to compare covariate-adjusted ROC curves for different markers.

ACKNOWLEDGEMENT

We thank Meir Stampfer for providing the data for the illustration. This paper was supported by funding from the U.S. National Institutes of Health.

APPENDIX

Proof of Theorem 1. We write

$$\begin{aligned}
 n_D^{1/2} \{ \text{AR}\hat{\text{OC}}_{\hat{\theta}}(t) - \text{AROC}_{\theta}(t) \} &= n_D^{1/2} \{ \text{AR}\hat{\text{OC}}_{\hat{\theta}}(t) - \text{AROC}_{\hat{\theta}}(t) \} \\
 &\quad + n_D^{1/2} \{ \text{AROC}_{\hat{\theta}}(t) - \text{AROC}_{\theta}(t) \} \\
 &= n_D^{-1/2} \sum_{i=1}^{n_D} I \{ Y_{D_i} > q(t; Z_{D_i}, \hat{\theta}) \} - \text{pr} \{ Y_D > q(t; Z_D, \hat{\theta}) \mid \hat{\theta} \} \\
 &\quad + n_D^{1/2} [\text{pr} \{ Y_D > q(t; Z_D, \hat{\theta}) \mid \hat{\theta} \} - \text{pr} \{ Y_D > q(t; Z_D, \theta) \}] \\
 &\equiv A_n + B_n.
 \end{aligned}$$

Note that $B_n = n_D^{1/2} \{ g(\hat{\theta}) - g(\theta) \}$, and by Assumptions 1–3, the delta method (Ferguson, 1996, p. 45) and Slutsky's theorem, B_n converges in distribution to a normal zero-mean random variable with variance matrix

$$\sigma_b^2 = \lambda \frac{\partial}{\partial \theta} \text{AROC}_{\theta}(t) \Sigma_{\theta} \frac{\partial}{\partial \theta^T} \text{AROC}_{\theta}(t).$$

Now, we write $A_n = n_D^{-1/2} \sum_{i=1}^{n_D} A_{n_i}$, and find its asymptotic distribution conditional on $\hat{\theta}$, using the Lindeberg–Feller central limit theorem. First, note that $E(A_{n_i} \mid \hat{\theta}) = 0$ and $\text{var}(A_{n_i} \mid \hat{\theta}) = \text{AROC}_{\hat{\theta}}(t)\{1 - \text{AROC}_{\hat{\theta}}(t)\}$. Convergence under the Lindeberg–Feller central limit theorem requires that

$$\frac{1}{n_D \text{AROC}_{\hat{\theta}}(t)\{1 - \text{AROC}_{\hat{\theta}}(t)\}} \sum_{i=1}^{n_D} E(A_{n_i}^2 I[|A_{n_i}| \geq \epsilon n_D \text{AROC}_{\hat{\theta}}(t)\{1 - \text{AROC}_{\hat{\theta}}(t)\}]) \quad (\text{A1})$$

converges to zero as $n_D \rightarrow \infty$ for all $\epsilon > 0$. However, $A_{n_i}^2 I[|A_{n_i}| \geq \epsilon n_D \text{AROC}_{\hat{\theta}}(t)\{1 - \text{AROC}_{\hat{\theta}}(t)\}]$ takes the value $\{1 - \text{AROC}_{\hat{\theta}}(t)\}^2 I\{\text{AROC}_{\hat{\theta}}(t)^{-1} \geq \epsilon n_D\}$ with probability $\text{AROC}_{\hat{\theta}}(t)$, and $\text{AROC}_{\hat{\theta}}(t)^2 I[\{1 - \text{AROC}_{\hat{\theta}}(t)\}^{-1} \geq \epsilon n_D]$ with probability $1 - \text{AROC}_{\hat{\theta}}(t)$. Hence, (A1) becomes

$$\{1 - \text{AROC}_{\hat{\theta}}(t)\} I\{\text{AROC}_{\hat{\theta}}(t)^{-1} \geq \epsilon n_D\} + \text{AROC}_{\hat{\theta}}(t) I[\{1 - \text{AROC}_{\hat{\theta}}(t)\}^{-1} \geq \epsilon n_D].$$

Assumptions 4 and 5 ensure that this converges to zero. Thus, conditional on $\hat{\theta}$, $A_n[\text{AROC}_{\hat{\theta}}(t)\{1 - \text{AROC}_{\hat{\theta}}(t)\}]^{-1/2}$ converges in distribution to a standard normal random variable as $n_D \rightarrow \infty$. Finally, the asymptotic distribution of $A_n[\text{AROC}_{\hat{\theta}}(t)\{1 - \text{AROC}_{\hat{\theta}}(t)\}]^{-1/2}$ conditional on $\hat{\theta}$ is the same as that of $A_n[\text{AROC}_{\hat{\theta}}(t)\{1 - \text{AROC}_{\hat{\theta}}(t)\}]^{-1/2}$ conditional on B_n , since it is functionally independent of $\hat{\theta}$ and $B_n = n_D^{1/2}\{g(\hat{\theta}) - g(\theta)\}$. By Assumptions 2 and 3, $\text{AROC}_{\hat{\theta}}(t)\{1 - \text{AROC}_{\hat{\theta}}(t)\}$ converges in probability to $\text{AROC}_{\theta}(t)\{1 - \text{AROC}_{\theta}(t)\}$ as $n_D \rightarrow \infty$. By Slutsky's theorem,

$$\begin{pmatrix} A_n[\text{AROC}_{\theta}(t)\{1 - \text{AROC}_{\theta}(t)\}]^{-1/2} \\ B_n \end{pmatrix},$$

converges in distribution to a zero-mean bivariate normal random variable with covariance matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & \sigma_b^2 \end{pmatrix},$$

as $n_D, n_{\bar{D}} \rightarrow \infty$. The continuous mapping theorem then yields the desired result. $\square$

Nonparametric estimation of $\text{AROC}(t)$

We prove that Assumptions 2 and 3 are satisfied with empirical quantile estimators. Under Assumption 6, by standard empirical process theory (Ferguson, 1996, p. 91), for a fixed stratum $Z = z$ and conditional on $n_{\bar{D}}(z), n_{\bar{D}}(z)^{1/2}\{\hat{S}_{\bar{D}}^{-1}(t \mid z) - S_{\bar{D}}^{-1}(t \mid z)\}$ converges in distribution to a normal zero-mean random variable with variance $\sigma_z^2 = t(1-t)f_{\bar{D}}^{-2}\{S_{\bar{D}}^{-1}(t \mid z) \mid z\}$. By Assumption 1, $n_{\bar{D}}(z)n_{\bar{D}}^{-1}$ converges in probability to $p_{\bar{D}}(z)$ as $n_{\bar{D}} \rightarrow \infty$. Hence, for all ϵ , there exists N such that

$$\begin{aligned} & \text{pr}[n_{\bar{D}}(z)^{1/2}\{\hat{S}_{\bar{D}}^{-1}(t \mid z) - S_{\bar{D}}^{-1}(t \mid z)\} \leq y] \\ &= E(\text{pr}[n_{\bar{D}}(z)^{1/2}\{\hat{S}_{\bar{D}}^{-1}(t \mid z) - S_{\bar{D}}^{-1}(t \mid z)\} \leq y \mid n_{\bar{D}}(z)]) \\ &= E(\text{pr}[n_{\bar{D}}(z)^{1/2}\{\hat{S}_{\bar{D}}^{-1}(t \mid z) - S_{\bar{D}}^{-1}(t \mid z)\} \leq y \mid n_{\bar{D}}(z)] I\{n_{\bar{D}}(z) > N\}) \\ &\quad + E(\text{pr}[n_{\bar{D}}(z)^{1/2}\{\hat{S}_{\bar{D}}^{-1}(t \mid z) - S_{\bar{D}}^{-1}(t \mid z)\} \leq y \mid n_{\bar{D}}(z)] I\{n_{\bar{D}}(z) \leq N\}) \\ &< \{\Phi(y/\sigma_z) + \epsilon\} \text{pr}\{n_{\bar{D}}(z) > N\} \\ &\quad + E(\text{pr}[n_{\bar{D}}(z)^{1/2}\{\hat{S}_{\bar{D}}^{-1}(t \mid z) - S_{\bar{D}}^{-1}(t \mid z)\} \leq y \mid n_{\bar{D}}(z)] I\{n_{\bar{D}}(z) \leq N\}). \end{aligned}$$

Since $n_{\bar{D}}(z) \rightarrow \infty$, the second term can be made arbitrarily small, and $\text{pr}\{n_{\bar{D}}(z) > N\}$ arbitrarily close to 1, by choosing N large enough. Thus, $n_{\bar{D}}(z)^{1/2}\{\hat{S}_{\bar{D}}^{-1}(t \mid z) - S_{\bar{D}}^{-1}(t \mid z)\}$ converges in distribution to a normal zero-mean random variable with variance σ_z^2 , and by Slutsky's theorem, $n_{\bar{D}}^{1/2}\{\hat{S}_{\bar{D}}^{-1}(t \mid z) - S_{\bar{D}}^{-1}(t \mid z)\}$ converges in distribution to a normal zero-mean random variable with variance $\sigma_z^2/p_{\bar{D}}(z)$ as $n_{\bar{D}} \rightarrow \infty$. Since observations in different strata are independent, marginal convergence implies joint asymptotic

normality, with the variance-covariance matrix $\Sigma_\theta = \text{diag}\{\sigma_z^2/p_{\bar{D}}(z)\}$. We also calculate the form of $(\partial/\partial\theta)\text{AROC}_\theta(t)$. We have

$$\begin{aligned}\frac{\partial}{\partial\theta_z}\text{AROC}_\theta(t) &= \frac{\partial}{\partial S_{\bar{D}}^{-1}(t|z)} E[\text{pr}\{Y_D > S_{\bar{D}}^{-1}(t|Z_D) | Z_D\}] \\ &= -E\left[f_D\{S_{\bar{D}}^{-1}(t|Z_D) | Z_D\} \frac{\partial}{\partial S_{\bar{D}}^{-1}(t|z)} S_{\bar{D}}^{-1}(t|Z_D)\right] \\ &= -f_D\{S_{\bar{D}}^{-1}(t|z) | z\} p_D(z),\end{aligned}$$

and $V(t)$ reduces to (3).

Consistency of the semiparametric variance estimator

We write the estimator of $(\partial/\partial\theta_j)\text{AROC}_\theta(t)$ as

$$\frac{\text{AROC}_{\hat{\theta}+h_j(n)}(t) - \text{AROC}_{\theta+h_j(n)}(t)}{2h(n)} + \frac{\text{AROC}_{\theta+h_j(n)}(t) - \text{AROC}_{\theta-h_j(n)}(t)}{2h(n)} - \frac{\text{AROC}_{\hat{\theta}-h_j(n)}(t) - \text{AROC}_{\theta-h_j(n)}(t)}{2h(n)}. \quad (\text{A2})$$

Consider the first component. We claim that $n_D^{1/2}\{\text{AROC}_{\hat{\theta}+h_j(n)}(t) - \text{AROC}_{\theta+h_j(n)}(t)\}$ converges in distribution to a normal zero-mean random variable with variance $V(t)$. The proof of this fact is similar to the proof that $n_D^{1/2}\{\text{AROC}_{\hat{\theta}}(t) - \text{AROC}_\theta(t)\}$ is asymptotically normal, and hence $O_p(1)$, as proven earlier. Hence, $n_D^{1/2}\{\text{AROC}_{\hat{\theta}+h_j(n)}(t) - \text{AROC}_{\theta+h_j(n)}(t)\}/\{2n_D^{1/2}h(n)\} \rightarrow 0$ in probability since the denominator converges to ∞ . A similar argument can be used to prove that the third term in (A2) converges to 0. Finally, $\{\text{AROC}_{\theta+h_j(n)}(t) - \text{AROC}_{\theta-h_j(n)}(t)\}/\{2h(n)\}$ converges in probability to $(\partial/\partial\theta_j)\text{AROC}_\theta(t)$ by continuity of $\text{AROC}_\theta(t)$ in θ ; see Assumption 3. Hence, our estimator of $(\partial/\partial\theta_j)\text{AROC}_\theta(t)$ is consistent. Now, with $\hat{\Sigma}_\theta$ converging in probability to Σ_θ , $\text{AROC}_{\hat{\theta}}(t)$ converging in probability to $\text{AROC}_\theta(t)$ and consistency of the derivative estimator, we have consistency of the composite variance estimator.

Consistency of the nonparametric variance estimator

We have $\hat{p}_D(z)$ converging in probability to $p_D(z)$ and $\hat{p}_{\bar{D}}(z)$ converging in probability to $p_{\bar{D}}(z)$ for $z = 1, \dots, K$, and, by standard empirical process theory (Ferguson, 1996, p. 91) and Assumption 6, $\hat{S}_{\bar{D}}^{-1}(t|z)$ converges in probability to $S_{\bar{D}}^{-1}(t|z)$ as $n_{\bar{D}} \rightarrow \infty$. We write

$$\begin{aligned}|\hat{f}_D\{\hat{S}_{\bar{D}}^{-1}(t|z) | z\} - f_D\{S_{\bar{D}}^{-1}(t|z) | z\}| &= |\hat{f}_D\{\hat{S}_{\bar{D}}^{-1}(t|z) | z\} - f_D\{\hat{S}_{\bar{D}}^{-1}(t|z) | z\} \\ &\quad + f_D\{\hat{S}_{\bar{D}}^{-1}(t|z) | z\} - f_D\{S_{\bar{D}}^{-1}(t|z) | z\}| \\ &\leq |\hat{f}_D\{\hat{S}_{\bar{D}}^{-1}(t|z) | z\} - f_D\{\hat{S}_{\bar{D}}^{-1}(t|z) | z\}| \\ &\quad + |f_D\{\hat{S}_{\bar{D}}^{-1}(t|z) | z\} - f_D\{S_{\bar{D}}^{-1}(t|z) | z\}|.\end{aligned}$$

The first term converges in probability to zero by the uniform consistency of $\hat{f}_D(y|z)$, while the second term converges in probability to zero by the consistency of $\hat{S}_{\bar{D}}^{-1}(t|z)$, Assumption 8 and the continuous mapping theorem. Hence, $\hat{f}_D\{\hat{S}_{\bar{D}}^{-1}(t|z) | z\}$ converges in probability to $f_D\{S_{\bar{D}}^{-1}(t|z) | z\}$ as $n_D, n_{\bar{D}} \rightarrow \infty$. A similar argument shows that $\hat{f}_{\bar{D}}\{\hat{S}_D^{-1}(t|z) | z\}$ is also consistent. The variance estimator is a continuous function of these components, and under Assumption 6 it is consistent.

REFERENCES

- ALONZO, T. A. & PEPE, M. S. (2002). Distribution-free ROC analysis using binary regression techniques. *Biostatistics* **3**, 421–32.

- BAKER, S. (2003). The central role of receiver operating characteristic (ROC) curves in evaluating tests for the early detection of cancer. *J. Nat. Cancer Inst.* **95**, 511–5.
- CAI, T. & PEPE, M. S. (2002). Semi-parametric ROC analysis to evaluate biomarkers for disease. *J. Am. Statist. Assoc.* **97**, 1099–107.
- COLE, T. J. (1990). The LMS method for constructing normalized growth standards. *Eur. J. Clin. Nutr.* **44**, 45–60.
- COLE, T. J. & GREEN, P. J. (1992). Smoothing reference centile curves: the LMS method and penalized likelihood. *Statist. Med.* **11**, 1305–19.
- ETZIONI, R., FALCON, S., GANN, P. H., KOOPERBERG, C. L., PENSON, D. F. & STAMPFER, M. J. (2004). Prostate-specific antigen and free prostate-specific antigen in the early detection of prostate cancer: do combination tests improve detection? *Cancer Epidemiol. Biomark. Prev.* **13**, 1640–5.
- FERGUSON, T. S. (1996). *A Course in Large Sample Theory*. London: Chapman and Hall.
- GANN, P. H., HENNEKENS, C. H. & STAMPFER, M. J. (1995). A prospective evaluation of plasma prostate-specific antigen for detection of prostatic cancer. *J. Am. Med. Assoc.* **273**, 289–94.
- HANLEY, J. A. (1988). The robustness of the ‘binormal’ assumptions used in fitting ROC curves. *Med. Decis. Making* **8**, 197–203.
- HANLEY, J. A. (1996). The use of the ‘binormal’ model for parametric ROC analysis of quantitative diagnostic tests. *Statist. Med.* **15**, 1575–85.
- HEAGERTY, P. & PEPE, M. S. (1999). Semi-parametric estimation of regression quantiles with application to standardizing weight for height and age in U.S. children. *Appl. Statist.* **48**, 533–51.
- JANES, H. & PEPE, M. S. (2008). Matching in studies of classification accuracy: implications for bias, efficiency, and assessment of incremental value. *Biometrics* **64**, 1–9.
- LIN, Y. & JEON, Y. (2003). Discriminant analysis through a semi-parametric model. *Biometrika* **90**, 379–92.
- PEPE, M. S. (2000). An interpretation for the ROC curve and inference using GLM procedures. *Biometrics* **56**, 352–9.
- PEPE, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford: Oxford University Press.
- PEPE, M. S. & CAI, T. (2004). The analysis of placement values for evaluating discriminatory measures. *Biometrics* **60**, 528–35.
- PEPE, M. S., ETZIONI, R., FENG, Z., POTTER, J. D., THOMPSON, M. L., THORNQUIST, M., WINGET, M. & YATSUI, Y. (2001). Phases of biomarker development for early detection of cancer. *J. Nat. Cancer Inst.* **93**, 1054–61.
- SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.
- SWETS, J. A. (1986). Indices of discrimination or diagnostic accuracy: their ROCs and implied methods. *Psychol. Bull.* **99**, 100–17.
- ZHENG Y. & HEAGERTY P. J. (2004). Semiparametric estimation of time-dependent ROC curves for longitudinal marker data. *Biostatistics* **5**, 615–32.

[Received May 2007. Revised June 2008]