Biostatistics (2013), **14**, 1, *pp.* 160–172
doi:10.1093/biostatistics/kxs020
Advance Access publication on June 20, 2012

ROC curve estimation under test-result-dependent sampling

XIAOFEI WANG*

Department of Biostatistics & Bioinformatics, Duke University Medical Center, Durham, NC 27710, USA

xiaofei.wang@duke.edu

JUNLING MA

Research Center for Applied Statistics, School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai 200433, China

STEPHEN L. GEORGE

Department of Biostatistics & Bioinformatics, Duke University Medical Center, Durham, NC 27710, USA

Summary

The receiver operating characteristic (ROC) curve is often used to evaluate the performance of a biomarker measured on continuous scale to predict the disease status or a clinical condition. Motivated by the need for novel study designs with better estimation efficiency and reduced study cost, we consider a biased sampling scheme that consists of a SRC and a supplemental TDC. Using this approach, investigators can oversample or undersample subjects falling into certain regions of the biomarker measure, yielding improved precision for the estimation of the ROC curve with a fixed sample size. Test-result-dependent sampling will introduce bias in estimating the predictive accuracy of the biomarker if standard ROC estimation methods are used. In this article, we discuss three approaches for analyzing data of a test-result-dependent structure with a special focus on the empirical likelihood method. We establish asymptotic properties of the empirical likelihood estimators for covariate-specific ROC curves and covariate-independent ROC curves and give their corresponding variance estimators. Simulation studies show that the empirical likelihood method yields good properties and is more efficient than alternative methods. Recommendations on number of regions, cutoff points, and subject allocation is made based on the simulation results. The proposed methods are illustrated with a data example based on an ongoing lung cancer clinical trial.

Keywords: Binormal model; Covariate-independent ROC curve; Covariate-specific ROC curve; Empirical likelihood method; Test-result-dependent sampling.

1. Introduction

Biomarkers have an important role in distinguishing diseased versus non-diseased subjects and in predicting patients with worse versus better prognosis or patients who are resistent versus sensitive to a

*To whom correspondence should be addressed.

treatment. Before a biomarker is able to be adopted for clinical use, validation studies are needed to assess how well it distinguishes disease conditions in a target population. The receiver operating characteristic (ROC) curve is an important tool for characterizing the performance of a diagnostic test or a biomarker in predicting disease conditions (Zhou, McClish, *and others*, 2002; Pepe, 2003). Let Y denote the test result of the biomarker measured on continuous scale. Let D denote the true disease condition with $D = 1$ for an unfavorable condition (e.g. diseased, short survival, resistent to treatment) and $D = 0$ for a favorable condition (e.g. non-diseased, long survival, sensitive to treatment). The true positive rate, $\mathrm{TPR}(c)$, and false positive rate, $\mathrm{FPR}(c)$, at threshold c are defined as $\Pr(Y \geqslant c|D = 1) \equiv S_1(c)$ and $\Pr(Y \geqslant c|D = 0) \equiv S_0(c)$, respectively. The ROC curve is a plot of the entire set of $\{(\mathrm{FPR}(c), \mathrm{TPR}(c)),\ c \in (-\infty, \infty)\}$, and it can be written as $\mathrm{ROC}(t) = S_1(S_0^{-1}(t))$, a function of $t = S_0(c)$ where $t \in (0, 1)$. In this article, we focus on the ROC curve for continuous biomarkers, though the proposed methods are applicable to test results or biomarkers measured on ordinal scale.

Rigorous evaluation of the accuracy of a biomarker predicting clinical conditions commonly requires a prospective study with enrollment of hundreds or thousands of subjects. With limited resources, there is a need for novel designs and statistical methods that allow investigators to conduct studies that require fewer patients, cost less and require less time to complete. Since the total study cost is often dominated by the cost spent on identifying the true disease conditions, an appealing idea is to limit the number of subjects whose true disease conditions will be identified. To achieve better efficiency, methods for sampling a cohort of subjects of a fixed number becomes crucial. Rather than selecting a simple random sample (SRS) from the target population, one can improve the efficiency of ROC estimation by oversampling patients from subpopulations that contain more information while undersampling patients with less information about the biomarker performance. When the prevalence rate of disease positivity in the target population is low, we may oversample those patients with high test results and undersample those with low or moderate test results. When we are particularly interested in the biomarker performance in certain regions of the test result, subjects with test results in the region of interest may be oversampled. When the biomarker performance in the entire range of the test result is of interest, subjects in low, middle, and high range of test results may be sampled with balanced allocation to maximize the estimation efficiency of the entire ROC curve. In this article, we consider a test-result-dependent sampling (TDS) design, which consists of two sampling components: a simple random component (SRC) and a test-result-dependent component (TDC). The SRC is a sample of subjects randomly selected from the target population while the TDC consists of multiple samples of subjects selected from strata defined by the ranges of the test result or the biomarker measure.

We use an ongoing phase III cancer clinical trial to illustrate the idea. COX2 expression is measured from lung cancer specimens with a range from 0 to 10 and higher COX2 score indicative of worse survival (Edelman *and others*, 2007). CALGB 30801 is a randomized clinical trial in COX2 positive non-small cell lung cancer patients. The study was designed to evaluate the survival benefit of celecoxib (a COX2 inhibitor) combined with standard chemotherapy over standard chemotherapy alone. A total of 216 COX2 positive patients will be randomized in a balanced 1:1 allocation to the two treatments. Patients with negative ($\mathrm{COX2} < 2$) and moderate ($2 \leqslant \mathrm{COX2} < 4$) expression will not receive celecoxib + standard chemotherapy since the agent is believed to have a detrimental effect on these patients. The prevalence rates of COX2 negatives, moderates and positives are expected to be approximately 60%, 13%, and 27%, respectively, in the target population. Therefore, approximately 800 patients are to be screened to accrue the 216 COX2 positives. As a secondary objective, the investigators are interested in validating the prognostic value of COX2 for survival among those patients who receive standard chemotherapy alone. Because the predictive performance of COX2 in the region from moderate to high scores is of primary interest, to avoid the costly option of treating the large number of COX2 negatives with standard chemotherapy and following them for long term survival, the investigators decided to select the first 200 patients treated on standard chemotherapy alone into the COX2 validation cohort. Based on the anticipated prevalence rates,

we expect 120 COX2 negatives, 26 COX2 moderates, and 54 COX2 positives from the 200 patients and these patients constitute the SRC of the TDS design. To further improve the efficiency of estimating the part of ROC curve in the range from moderate to high COX2 scores, the rest of 78 COX2 moderates and 54 COX2 positives treated by standard chemotherapy alone at the study closure forms the TDC.

Whenever the ascertained data consist of a sampling component other than an SRS, standard ROC methods assuming an SRS will yield biased estimates. Our goal is to investigate the utility of a TDS design in biomarker validation and develop appropriate ROC estimation methods for the data arising from such a design. For data arising from an SRS design, non-parametric and parametric procedures have been proposed for ROC curve estimation (Dorfman and Alf, 1969; Metz, 1978). To understand the influence of covariates on the accuracy of a diagnostic test, investigators are increasingly interested in the *covariate-specific* ROC curve, such as the ROC curve for patient subgroups with specified covariate profiles (e.g. female and age >65). Two regression approaches for covariate-specific ROC curve have been proposed in the literature. The first approach specifies a model for the test result as a function of disease conditions and covariates. Covariate effects on the ROC curve can then be calculated by deriving the induced form of ROC curve (Tosteson and Begg, 1988; Toledano and Gatsonis, 1995). A second approach, proposed by Pepe (1997), directly models the ROC curve itself using the model $\phi(\mathrm{ROC}_X(t)) = \psi(t) + \beta X$, where $\phi(\cdot)$ and $\psi(\cdot)$ denote monotonic increasing functions on $(0, 1)$. For data arising from a TDS design, we propose a modified inverse probability weighting (IPW) method, a non-parametric method and an empirical likelihood method. The empirical likelihood method will be emphasized because of its potential for increased efficiency. We adopt the binormal model formulated by Pepe (2003), which works nicely with the empirical likelihood method. The binormal model characterizes the relationship between the test result Y and the disease status D as well as the effects of covariates X and their interactions with D. The covariate-specific ROC curve is induced by evaluating at the estimates of model parameters. By use of empirical probability estimates for covariates, we can average over the effects of covariates to obtain *covariate-independent* ROC curve. In the context of biased sampling, similar approaches were studied by Zhou, Weaver, *and others* (2002) and Wang and Zhou (2006). Qin *and others* (2009) studied a unified empirical likelihood approach to the problem of missing data and their method achieves semiparametric efficiency if missingness mechanism is correctly specified. Their method can be extended to analyze data arising from the TDS design if the sampling probability of subjects is identifiable.

Verification bias is a well-known research area in the literature of diagnostic medicine (see Zhou, Weaver, *and others*, 2002; Pepe, 2003). The TDS design is related to the verification bias sampling in that both sampling schemes introduce bias when standard ROC methods are used in the analysis. However, there are important differences between them, which make the existing ROC methods for verification bias not directly applicable to TDS data. Verification bias occurs primarily in population-based screening studies, in which a screening test is applied to all N subjects in the cohort, yielding a test result Y for all study units. As only a subset of patients is selected to observe (verify) true disease status D, unselected patients in the cohort are missing D. When D is missing at random (Little and Rubin, 1987), the selection probability of observing D is estimable. Statistical methods were proposed by weighing each verified subject inversely with the selection probability to correct the bias (e.g. Begg and Greenes, 1983; Gray *and others*, 1984; Alonzo and Pepe, 2005). In the TDS design, patients are selected to observe D depending on the test result Y, but the size of the parent cohort from which the patients are accrued and the test results of unscreened patients are unknown to the investigators. We assume that the test results of the patients who are screened but unselected for observing true disease condition are not available to the investigators. In many hospital-based studies and randomized clinical trials, the test results for preregistered but unselected patients will not be captured and stored electronically. For all the subjects in the selected cohort, the data available for analysis are observed, both Y and D. As there are no missing variables, the IPW method is not applicable to the TDS data analysis. In some cases such as the COX2 study, however, the investigators may know the test results of all patients screened for the trial. Using the screened

patients as the parent cohort, we are able to apply the existing methods for verification bias to such TDS data. Since the size of unselected patients is often small compared with the size of the target population, the extent of efficiency gain by utilizing the extra information on test results with the verification bias methods remains unknown. We will investigate this issue via simulation. Nevertheless, the estimation of a covariate-specific ROC curve has not been a focus of existing methods for verification bias, and our proposed TDS methods are useful since it requires no information on unselected patients and estimates both covariate-specific and covariate-independent ROC curves.

The rest of the paper is organized as follows. Section 2 reviews the covariate-specific ROC curve and the covariate-independent ROC curve under a binormal model. In Section 3, we specify the data structure and the likelihood function for the TDS design. In Section 4, we propose an empirical likelihood-based method for the covariate-specific ROC curve and the covariate-independent ROC curve, and establish the asymptotic properties of the proposed estimators. Section 5 presents two alternative methods for data arising from the TDS design: the weighted likelihood method and a non-parametric method. Section 6 evaluates finite sample properties of the proposed methods for ROC curves, compares the relative efficiency of these methods under the TDS design, and compares the efficiency of the TDS design relative to the SRS design. We illustrate in Section 7 the proposed TDS methods using data from the COX2 study described above. Discussion is given in Section 8. Proofs of the asymptotic results in Section 4 are given in the supplementary material (available at *Biostatistics* online). Additional simulation results on the comparison of three TDS methods, the impact of choices of number of regions, cutoff points and subject allocation as well as the robustness of binormal model are also left to the supplementary material (available at *Biostatistics* online).

2. Induced covariate-specific and covariate-independent ROC curves

To evaluate covariate-specific ROC curve, we adopt the approach of Tosteson and Begg (1988) and Toledano and Gatsonis (1995) to induce the ROC curve from the estimates of an underlying parametric regression model. Recall that Y is the continuous test result, D is the true disease condition, and X is the vector of covariates that affects the test accuracy. We adopt a binormal ROC model (see Pepe, 2003) to characterize the relationship between Y and (D, X):

$$Y = \beta_0 + \beta_D D + \beta_X{}^{\mathrm{T}} X + \beta_{DX}{}^{\mathrm{T}} D X_D + \sigma(D)\epsilon, \tag{2.1}$$

where $\epsilon \sim N(0, 1)$ and $\sigma(D) = \sigma_1 I[D = 1] + \sigma_0 I[D = 0]$. DX_D is the interaction term between D and X, where X_D could be X or a subset of X. This formulation allows the effects of X on Y and the variance of Y to differ for diseased and non-diseased subjects. The binormal ROC model is invariant to monotonic increasing transformation on test result Y.

Let $S_{1X}(c) = \Pr(Y \geqslant c|X, D = 1)$ and $S_{0X}(c) = \Pr(Y \geqslant c|X, D = 0)$. For any given vector of covariates X, we can write the covariate-specific ROC curve as $\mathrm{ROC}_X(t) = S_{1X}(S_{0X}^{-1}(t))$. Under model (2.1), we have $S_{1X}(c) = \Phi(\sigma_1^{-1}(\beta_0 + \beta_D + \beta_X^{\mathrm{T}} X + \beta_{DX}^{\mathrm{T}} X_D - c))$ and $S_{0X}(c) = \Phi(\sigma_0^{-1}(\beta_0 + \beta_X^{\mathrm{T}} X) - c)$. The induced covariate-specific ROC curve is given by

$$\mathrm{ROC}_X(t) = S_{1X}(S_{0X}^{-1}(t)) = \Phi\left(\frac{\beta_D + \beta_{DX}^{\mathrm{T}} X_D + \sigma_0 \Phi^{-1}(t)}{\sigma_1}\right) \tag{2.2}$$

corresponding to a binormal ROC curve with an intercept $a = (\beta_D + \beta_{DX}^{\mathrm{T}} X_D)/\sigma_1$ and a slope $b = \sigma_0/\sigma_1$. Only those covariates included in X_D have effect on the shape of the ROC curve. Furthermore, we can obtain a covariate-independent ROC curve by averaging $S_{1X}(c)$ and $S_{0X}(c)$ over the distribution of X. For any $c = S_0^{-1}(t)$, the covariate-independent ROC curve $\mathrm{ROC}(t) = S_1(S_0^{-1}(t))$ where $S_1(c) = \int S_{1X}(c)\, \mathrm{d}G(X|D = 1)$, $S_0(c) = \int S_{0X}(c)\, \mathrm{d}G(X|D = 0)$ and $G(X|D)$ is the conditional cumulative distribution of X on D.

3. DATA STRUCTURE AND LIKELIHOOD

Suppose the test result Y falls into one of K mutually exclusive regions, denoted by $C_k = (a_{k-1}, a_k)$, $k = 1, 2, \ldots, K$ with $a_{k-1} < a_k$, $a_0 = -\infty$ and $a_K = \infty$. Under the TDS design, there are two sampling components: an SRC and a TDC. In the SRC, the observed data are denoted by $\{Y_{0i}, D_{0i}, X_{0i}\}$ with $i = 1, 2, \ldots, n_0$. In the TDC, one observes X and D conditional on the stratum C_k that Y falls into so that the data can be denoted by $\{Y_{ki}, D_{ki}, X_{ki} | Y \in C_k\}$ for $i = 1, 2, \ldots, n_k$. The total sample size with the two sampling components combined is $n = \sum_{s=0}^{K} n_s$.

Denote by $f(Y, D, X)$ the joint density function of (Y, D, X) and by $f_k(Y, D, X | Y \in C_k)$ the joint density function of (Y, D, X) given $\{Y \in C_k\}$. The likelihood function of the observed data is $L = [\prod_{i=1}^{n_0} f(Y_{0i}, D_{0i}, X_{0i})][\prod_{k=1}^{K} \prod_{i=1}^{n_k} f_k(Y_{ki}, D_{ki}, X_{ki} | Y_{ki} \in C_k)]$. The part in the first bracket is the data contributed by the SRC and the second bracket is that of the TDC. The observed likelihood is rewritten as

$$L(\beta) = \left[\prod_{i=1}^{n_0} h_\beta(Y_{0i} | D_{0i}, X_{0i}) g(D_{0i}, X_{0i})\right] \left[\prod_{k=1}^{K} \prod_{i=1}^{n_k} f_k(Y_{ki}, D_{ki}, X_{ki} | Y_{ki} \in C_k)\right], \tag{3.1}$$

where $g(D, X)$ is the joint density function of (D, X), $h_\beta(Y | D, X)$ is the conditional density function of Y on $\{D, X\}$, and $\boldsymbol{\beta} = (\beta_0, \beta_D, \beta_X^{\mathrm{T}}, \beta_{DX}^{\mathrm{T}}, \sigma_0, \sigma_1)^{\mathrm{T}}$ is the vector of regression coefficients and scale parameters in model (2.1). By Bayes theorem, we have $f_k(Y_{ki}, D_{ki}, X_{ki} | Y_{ki} \in C_k) = I[Y_{ki} \in C_k] h_\beta(Y_{ki} | D_{ki}, X_{ki}) g(D_{ki}, X_{ki}) \Pr^{-1}(Y_{ki} \in C_k)$, where $\Pr(Y_{ki} \in C_k) = \int [\int_{C_k} h_\beta(y | d, x) \mathrm{d}y] \mathrm{d}G(d, x)$, and $G(D, X)$ is the joint distribution function of (D, X). Let $\theta_k = \Pr(Y \in C_k)$ and $\eta = (\beta^{\mathrm{T}}, \theta_1, \ldots, \theta_{K-1})^{\mathrm{T}}$. Writing (3.1) as a function of η and $G(D, X)$, we have

$$L(\eta, G) = \left[\prod_{s=0}^{K} \prod_{i=1}^{n_s} h_\beta(Y_{si} | D_{si}, X_{si})\right] \left[\prod_{s=0}^{K} \prod_{i=1}^{n_s} g(D_{si}, X_{si})\right] \prod_{k=1}^{K} \theta_k^{-n_k}, \tag{3.2}$$

where $\theta_K = 1 - \sum_{k=1}^{K-1} \theta_k$. In practice, $G(D, X)$ is difficult to be estimated using parametric modeling due to the potential high-dimensional nature of (D, X) and mis-specification of the distribution could lead to biased estimation on η. Statistical approach that does not rely on a parametric modeling on G is preferred.

4. ROC CURVE ESTIMATION UNDER TDS DESIGN

4.1 *Estimation of ROC model parameters*

In the context of biased sampling, Zhou, Weaver, *and others* (2002) and Wang and Zhou (2006) developed a profiled empirical likelihood approach that estimates model parameters by maximizing $L(\eta, G(\cdot))$ without parametric modeling on $G(\cdot)$. We adopt this approach to estimate parameters in model (2.1) under the TDS design. The idea is that for fixed β we can estimate $G(\cdot)$ non-parametrically by solving a set of estimation equations with constraints. By plugging in the estimate of $G(\cdot)$, we can rewrite $L(\eta, G(\cdot))$ as a form of profile likelihood on which the maximization with respect to the model parameters can be carried out. Let $p_{si} = g(d_{si}, x_{si})$ and rewrite the log of (3.2) as

$$l(\eta, \{p_{si}\}) = \sum_{s=0}^{K} \sum_{i=1}^{n_s} \log p_{si} - \sum_{k=1}^{K} (n_k \log \theta_k) + \sum_{s=0}^{K} \sum_{i=1}^{n_s} \log h_\beta(y_{si} | d_{si}, x_{si}), \tag{4.1}$$

where $h_\beta(y_{si} | d_{si}, x_{si})$ is a normal density function under the binormal assumption. Based on the empirical likelihood theory (e.g., Qin and Lawless, 1994; Owen, 2001), to estimate p_{si}, it is sufficient to search the

discrete probability space defined by the observed values of $\{D, X\}$ under the following constraints

$$\left\{ \sum_{s=0}^{K} \sum_{i=1}^{n_s} p_{si} = 1, \quad \sum_{s=0}^{K} \sum_{i=1}^{n_s} p_{si}(\theta_k - \Pr(y_{si} \in C_k | d_{si}, x_{si})) = 0, k = 1, \ldots, K-1 \right\},$$

where $\Pr(y_{si} \in C_k | d_{si}, x_{si}) = \int_{C_k} h_\beta(y | d_{si}, x_{si})\, \mathrm{d}y$. Using a Lagrange multiplier argument, we can derive the $\tilde{p}_{si}$ over which $l(\eta, \{p_{si}\})$ attains the maximum:

$$\tilde{p}_{si} = \frac{1}{n_0 + \sum_{k=1}^{K} (n_k/\theta_k) \Pr(y_{si} \in C_k | d_{si}, x_{si})}. \tag{4.2}$$

Now we plug $\tilde{p}_{si}$ into (4.1) and obtain a profile likelihood function

$$\begin{aligned} l_p(\eta) = &- \sum_{s=0}^{K} \sum_{i=1}^{n_s} \log \left(n_0 + \sum_{k=1}^{K} \frac{n_k}{\theta_k} \Pr(y_{si} \in C_k | d_{si}, x_{si}) \right) \\ &- \sum_{k=1}^{K} (n_k \log \theta_k) + \sum_{s=0}^{K} \sum_{i=1}^{n_s} \log h_\beta(y_{si} | d_{si}, x_{si}). \end{aligned} \tag{4.3}$$

The estimate $\hat{\eta} = (\hat{\beta}^{\mathrm{T}}, \hat{\theta}^{\mathrm{T}})^{\mathrm{T}}$ is obtained by solving the score equation $\partial l_p(\eta)/\partial \eta = 0$ using an iterative algorithm. Using the general theory of empirical likelihood, we can show under regularity conditions $n^{1/2}(\hat{\eta} - \eta)$ converges to a mean zero normal distribution in a neighborhood of η, and $n^{1/2}(\hat{G}(D, X) - G(D, X))$ converges to a mean zero Gaussian process on $\mathcal{D} \times \mathcal{X}$.

4.2 *Estimation of covariate-specific ROC curve*

We can induce the covariate-specific $\mathrm{ROC}_X(t)$ curve using the empirical likelihood estimate $\hat{\beta}$. For any given threshold c and any given vector of covariates X, we estimate $S_{1X}(c)$ and $S_{0X}(c)$ by $\hat{S}_{1X}(c) = \Phi((\hat{\mu}_{1X} - c)/\hat{\sigma}_1)$ and $\hat{S}_{0X}(c) = \Phi((\hat{\mu}_{0X} - c)/\hat{\sigma}_0)$, where $\hat{\mu}_{1X} = \hat{\beta}_0 + \hat{\beta}_D + \hat{\beta}_X^{\mathrm{T}} X + \hat{\beta}_{DX}^{\mathrm{T}} X_D$ and $\hat{\mu}_{0X} = \hat{\beta}_0 + \hat{\beta}_X^{\mathrm{T}} X$. The $\mathrm{ROC}_X(t)$ is estimated by $\widehat{\mathrm{ROC}}_X(t) = \hat{S}_{1X}(\hat{S}_{0X}^{-1}(t))$ and the curve itself is generated by plotting $\hat{S}_{1X}(c)$ against $\hat{S}_{0X}(c)$ for all c at the given values of X.

Let n_0^D be the number of diseased subjects of the SRC and n_k^D the number of diseased subjects in the kth stratum of Y, $k = 1, \ldots, K$ of the TDC. The first derivative of the profile likelihood function $l_p(\eta)$ with respect to β has a form of summation, denoted as $\partial l_p(\eta)/\partial \beta = \sum_{s=0}^{K} \sum_{i=1}^{n_s} \mathrm{grad}'_{si}$. Let $\varpi_{si}^{(k)} = \mathrm{grad}'_{si} I(d_{si} = k)$, $k = 0, 1$, and $H_\beta = \lim_{n \to \infty} (1/n)(\partial^2 l_p(\eta)/\partial \beta \partial \beta^{\mathrm{T}})$. Denote by $(\partial S_{kX}(c)/\partial \beta)$ and $\partial S_{kX}(v_t)/\partial v_t$, $k = 0, 1$ the derivatives of $S_{kX}(v_t)$ with respect to β and v_t, respectively. We have the following asymptotic property for the estimator $\widehat{\mathrm{ROC}}_X(t)$.

THEOREM 1 Assume $n_s^D/n_s \to \rho_s^D$, $s = 0, 1, \ldots, K$ with $0 < \rho_0^D < 1$ and $0 \leqslant \rho_1^D, \ldots, \rho_K^D < 1$ as $n \to \infty$. Under the conditions specified in Lemma 1, we have $n^{1/2}(\widehat{\mathrm{ROC}}_X(t) - \mathrm{ROC}_X(t))$ converges to a mean zero Gaussian process on $(0, 1)$ with variance $\Omega_1 + \Omega_0$ for any given t, where $\Omega_1 = [\partial S_{1X}(c)/\partial \beta^{\mathrm{T}}] H_\beta^{-1} \sum_{s=0}^{K} \rho_s \rho_s^D \mathrm{var}(\varpi_{si}^{(1)}) H_\beta^{-1} [\partial S_{1X}(c)/\partial \beta]$, and

$$\Omega_0 = \left[\left(\frac{\partial S_{0X}(v_t)}{\partial v_t} \right)^{-1} \frac{\partial S_{1X}(v_t)}{\partial v_t} \right]^2 \left[\frac{\partial S_{0X}(c)}{\partial \beta^{\mathrm{T}}} \right] H_\beta^{-1} \sum_{s=0}^{K} \rho_s (1 - \rho_s^D) \mathrm{var}(\varpi_{si}^{(0)}) H_\beta^{-1} \left[\frac{\partial S_{0X}(c)}{\partial \beta} \right].$$

4.3 *Estimation of covariate-independent ROC curve*

Notice that $\hat{G}(d,x) = 1/n \sum_{s=0}^{K} \sum_{i=1}^{n_s} \hat{p}_{si} I(x_{si} \leqslant x, d_{si} = d)$, $d = \{0,1\}$, where $\hat{p}_{si}$ is the estimate of $\tilde{p}_{si}$ by substituting $\hat{\eta}$ for η. We are able to estimate the covariate-independent ROC curve by averaging $\hat{S}_{1X}(c)$ and $\hat{S}_{0X}(c)$ over the empirical distribution $\hat{G}(d,x)$. The estimators $\hat{S}_1(c)$ and $\hat{S}_0(c)$ are given by

$$\hat{S}_1(c) = \frac{\int \hat{S}_{1X}(c)\, \mathrm{d}\hat{G}(D=1,x)}{\int \mathrm{d}\hat{G}(D=1,x)} = \frac{\sum_{s=0}^{K} \sum_{i=1}^{n_s} \hat{p}_{si} I(d_{si}=1) \hat{S}_{1X_{si}}(c)}{\sum_{s=0}^{K} \sum_{i=1}^{n_s} \hat{p}_{si} I(d_{si}=1)}, \tag{4.4}$$

$$\hat{S}_0(c) = \frac{\int \hat{S}_{0X}(c)\, \mathrm{d}\hat{G}(D=0,x)}{\int \mathrm{d}\hat{G}(D=0,x)} = \frac{\sum_{s=0}^{K} \sum_{i=1}^{n_s} \hat{p}_{si} I(d_{si}=0) \hat{S}_{0X_{si}}(c)}{\sum_{s=0}^{K} \sum_{i=1}^{n_s} \hat{p}_{si} I(d_{si}=0)}. \tag{4.5}$$

The estimator of the covariate-independent ROC is given by $\hat{\mathrm{ROC}}(t) = \hat{S}_1(\hat{S}_0^{-1}(t))$.

For convenience, let $G_1 = G(D=1,X)$, $\tilde{g}_1 = (1/(\rho_0 + \sum_{k=1}^{K}(\rho_k/\theta_k)\Pr(Y \in C_k|d,x)))I(d=1)$, $P_{si} = (\Pr(y_{si} \in C_1|d_{si},x_{si}) - \theta_1, \ldots, \Pr(y_{si} \in C_{K-1}|d_{si},x_{si}) - \theta_{K-1})^{\mathrm{T}}$, $S = \lim_{n\to\infty} 1/n \sum_{s=0}^{K} \sum_{i=1}^{n_s} P_{si} P_{si}^{\mathrm{T}}$, $B = \lim_{n\to\infty} 1/n \sum_{s=0}^{K} \sum_{i=1}^{n_s} P_{si}$, $\partial l_p(\eta)/\partial\eta = \sum_{s=0}^{K} \sum_{i=1}^{n_s} \mathrm{grad}_{si}$, $H_\eta = \lim_{n\to\infty}(1/n)(\partial^2 l_p(\eta)/\partial\eta\partial\eta^{\mathrm{T}})$, and $\zeta_{si} = 1 - G(d_{si},x_{si}) - (S^{-1}B)^{\mathrm{T}} P_{si}$. Define

$$\begin{aligned}\omega_{si}^{(1)} = {} & \left(\int \mathrm{d}G_1\right)^{-1} \left[\int \frac{\partial S_{1X}(c)}{\partial\eta}\, \mathrm{d}G_1 + \int S_{1X}(c)(\partial\tilde{g}_1/\partial\eta)\, \mathrm{d}x\right] (-\gamma^{-1}H_\eta)^{-1}\, \mathrm{grad}_{si}\, I(d_{si}=1) \\ & + \left(\int \mathrm{d}G_1\right)^{-1} [S_{1X_{si}}(c) - S_1(c)]\zeta_{si} I(d_{si}=1),\end{aligned} \tag{4.6}$$

where $\gamma = \sum_{s=0}^{K} \rho_s \rho_s^D$, $\omega_{si}^{(0)}$ is defined similar to $\omega_{si}^{(1)}$ for $D=0$. Asymptotic property for $\hat{\mathrm{ROC}}(t)$ is summarized in Theorem 2.

THEOREM 2 Under the condition of Theorem 1, we have $n^{1/2}(\hat{\mathrm{ROC}}(t) - \mathrm{ROC}(t))$ converges to a mean zero Gaussian process on $(0,1)$ with variance $\Gamma_1 + \Gamma_2$ for any given t, where $\Gamma_1 = \sum_{s=0}^{K} \rho_s \rho_s^D \mathrm{var}(\omega_{si}^{(1)})$ and $\Gamma_2 = [(\partial S_0(v_t)/\partial v_t)^{-1}(\partial S_1(v_t)/\partial v_t)]^2 \sum_{s=0}^{K} \rho_s(1-\rho_s^D)\, \mathrm{var}(\omega_{si}^{(0)})$.

5. ALTERNATIVE METHODS FOR ROC CURVE ESTIMATION UNDER TDS

Two alternative methods, a weighted likelihood method and a non-parametric method, for the TDS design are also investigated. The weighted likelihood method can be considered as a modified IPW method. Under the TDS design, we partition the SRC into K strata using the same cutoff points $(a_0, \ldots, a_K)$ for the TDC. We can write the pooled data as $\{Y_{ki}, D_{ki}, X_{ki}, k = 1, \ldots, K, i = 1, \ldots, n_{0,k} + n_k\}$, where $n_{0,k}$ is the size of the kth stratum in the SRC with $\sum_{k=1}^{K} n_{0,k} = n_0$. We estimate the weight for the ith subject in the kth stratum as $\tilde{\pi}_{ki} = (n_{0,k}/n_0(n_{0,k}+n_k))I(Y_{ki} \in C_k)$. This is indeed the same estimator proposed by Morgenthaler and Vardi (1986) for general biased sampling problem. The conditional likelihood function with weights $\tilde{\pi}_{ki}$ under model (2.1) for the TDS data has the form $L_{WL}(\beta) = \sum_{k=1}^{K} \sum_{i=1}^{n_{0,k}+n_k} \{\tilde{\pi}_{ki} \log f(Y_{ki}|D_{ki},X_{ki})\}$. The estimate $\hat{\beta}_{WL}$ is obtained by maximizing $L_{WL}(\beta)$. Following the same approach for the empirical likelihood method and replacing $\hat{p}_{si}$ with $\tilde{\pi}_{ki}$, we can obtain the estimates for $\mathrm{ROC}_X(t)$ and $\mathrm{ROC}(t)$. Since the weight $\tilde{\pi}_{ki}$ is estimated by proportions of patients in each stratum and does not depend on the ROC model, the weighted likelihood method is not a fully likelihood-based method and we expect that it is less efficient than the empirical likelihood method discussed in Section 4.

The non-parametric method makes no parametric assumption on the relationship between test result and disease status. Denote by $f(Y, D)$ the joint density of (Y, D), and $f_k(Y, D|Y \in C_k)$ is the joint density given $\{Y \in C_k\}$. We have the following likelihood function $L_{NP}(\cdot) \propto [\prod_{i=1}^{n_0} f(Y_{0i}, D_{0i})][\prod_{k=1}^{K} \prod_{i=1}^{n_k} f_k(Y_{ki}, D_{ki} \mid Y_{ki} \in C_k)]$. By maximizing $L_{NP}(\cdot)$ under similar constraints to those for (4.1), we can estimate $f(Y_{si}, D_{si})$. The covariate-independent ROC curve is obtained empirically by averaging the corresponding observations with the estimates of $f(Y_{si}, D_{si})$. Wang *and others* (2012) studied a similar non-parametric estimation method for the area under ROC curve (AUC) and partial AUC. However, this approach is not applicable to estimate the covariate-specific ROC curve. Because no information from covariates X is utilized, we expect the non-parametric method will be less efficient than the other two TDS methods.

6. Simulation

We investigated the finite sample properties of the proposed methods for ROC curve estimation under the TDS design and their relative efficiencies via a series of simulation studies. We also studied the possible efficiency gain of the TDS design over the SRS design. The findings are summarized in this section. Other issues, such as the comparison of three TDS methods, the impact of choice of number of regions, cutoff points and subject allocation on the precision of ROC estimation, and the robustness of the binormal ROC model, were also investigated. The details of these simulation results can be found in the supplementary material (available at *Biostatistics* online).

In all simulation studies, we generated data according to the model $Y = \beta_0 + \beta_D D + \beta_X X + \beta_{DX} DX + \sigma(D)\epsilon$, where $\epsilon \sim N(0, \sigma_D^2)$, and $\sigma_D^2 = I(D = 1)\sigma_1^2 + I(D = 0)\sigma_0^2$, $D \sim Bernoulli(0.3)$, $X \sim N(0, 1.2^2)$. We fixed $\beta_0 = 0.5$, $\beta_D = 1.0$, $\beta_X = 0.5$, $\beta_{DX} = 0.5$, $\sigma_1 = 1.2$, and $\sigma_0 = 1.0$ unless specified otherwise. In each independent run, we first generated a random sample of size n_0 and then generated the TDC of size n_{TDC}. Unless stated otherwise, the TDC subjects are from three regions—the lower region, the middle and the upper region of Y—with sizes of n_1, n_2, n_3, respectively. The following specification on subject allocation $(n_0, n_1, n_2, n_3) = (150, 50, 50, 50)$ and cutoff points $(a_1, a_2) = (\mu_Y - \sigma_Y, \mu_Y + \sigma_Y)$ were used. All simulation studies were based on 5000 independent runs.

6.1 *Performance of empirical likelihood method SL_{TDS}*

As noted in the Section 1, using the standard methods developed for SRS to estimate ROC curve on data arising from the TDS design will lead to biased results. To illustrate this, we applied the standard likelihood-based ROC estimation method PE_{SRS} to the TDS data. Figure S1 in the supplementary material (available at *Biostatistics* online) demonstrates the bias in ROC curve estimates from PE_{SRS} on a single simulated TDS dataset. In this specific case, PE_{SRS} underestimates the covariate-specific ROC curve $\text{ROC}_X(t)$ but overestimates the covariate-independent ROC curve $\text{ROC}(t)$. We further investigated the finite sample performance of the SL_{TDS} under the TDS design. At false positive rates $t = (0.1, 0.3, 0.5, 0.7, 0.9)$, Table 1 lists the mean of $\text{ROC}_X(t)$ and $\text{ROC}(t)$ (Estimate), the percent bias relative to true values (Bias%), the simulated standard error (SE) and the estimated standard error ($\widehat{\text{SE}}$), and the coverage probability of 95% confidence interval (95% CP). We used the variance estimators given in Section 4 to compute the $\widehat{\text{SE}}$ and the 95% CP. The relative biases associated with the proposed estimators are small for both covariate-specific $\text{ROC}_X(t)$ and covariate-independent $\text{ROC}(t)$; the estimated standard error at each t is close to its simulated standard error; and the coverage probability of the 95% confidence interval is close to its nominal level except for $\widehat{\text{ROC}}_X(t)$ as t is close to 1.

6.2 *Efficiency comparison of TDS and SRS designs*

The TDS design is expected to gain efficiency over the SRS design by oversampling subjects in "tailed" regions of the test result. We would like to confirm this via simulation. In each run of simulation, a TDS

Table 1. *ROC curve estimated by SL_{TDS} under TDS*

		True	Estimate	Bias%	SE	$\widehat{\mathrm{SE}}$	95% CP
$\hat{\mathrm{ROC}}_X(t)$							
$X=0$	$t=0.1$	0.407	0.412	1.23	0.0468	0.0479	0.954
	$t=0.3$	0.654	0.658	0.61	0.0425	0.0465	0.967
	$t=0.5$	0.798	0.800	0.25	0.0347	0.0378	0.957
	$t=0.7$	0.898	0.899	0.11	0.0248	0.0261	0.946
	$t=0.9$	0.971	0.971	0.00	0.0114	0.0113	0.918
$X=1.2$	$t=0.1$	0.605	0.607	0.33	0.0620	0.0666	0.961
	$t=0.3$	0.815	0.815	0.00	0.0443	0.0502	0.966
	$t=0.5$	0.909	0.907	−0.22	0.0293	0.0333	0.957
	$t=0.7$	0.962	0.960	−0.21	0.0166	0.0186	0.945
	$t=0.9$	0.992	0.991	−0.10	0.0055	0.0059	0.916
$\hat{\mathrm{ROC}}(t)$							
	$t=0.1$	0.385	0.388	1.04	0.0429	0.0436	0.938
	$t=0.3$	0.588	0.593	0.85	0.0416	0.0420	0.930
	$t=0.5$	0.724	0.724	0.00	0.0373	0.0381	0.940
	$t=0.7$	0.829	0.830	0.12	0.0309	0.0327	0.941
	$t=0.9$	0.930	0.929	−0.11	0.0198	0.0228	0.949

dataset with total size of $n=n_0+n_1+n_2+n_3=300$ was generated and an SRS dataset of the same size 300 was independently drawn. The empirical likelihood method SL_{TDS} was applied to the TDS dataset, while the standard method PE^*_{SRS} was applied to the SRS dataset. These two methods are chosen because they are both likelihood-based method and are the most efficient for the corresponding design. Here we use PE^*_{SRS} to denote the case that the standard ROC method is applied to an independently generated SRS sample, different from PE_{SRS} in Figure S1, where the standard ROC method was applied to the same TDS dataset that SL_{TDS} was applied to. Table 2 lists the ratio of $\mathrm{MSE}(SL_{\mathrm{TDS}})$ under the TDS design to that of $\mathrm{MSE}(PE^*_{\mathrm{SRS}})$ under the SRS design, where the sizes of the two designs are both $n=300$. It can be seen that SL_{TDS} yields smaller MSEs than PE^*_{SRS} under different specifications of subject allocations and cutoff points. We conclude that the TDS design is more efficient than the SRS design of the same size.

6.3 *Comparison with verification bias design and method*

In some TDS studies, the test results of screened but unselected patients are also known to investigators. The methods for verification bias are applicable when the screened patients are considered as the parent cohort. We conducted simulation to compare the empirical likelihood method SL_{TDS} and the weighted likelihood method WL_{TDS} with verification bias methods. Since none of the existing methods of verification bias is focused on covariate-specific ROC curve estimation, we propose an IPW method IPW_{VB} with the selection probability $\pi_k=(n_{0k}+n_k)/N_k$, where n_{0k}, n_k, and N_k are the sizes of the kth stratum for the SRC, the TDC, and the entire cohort, respectively. We started to generate a parent cohort of $N=1000$ subjects with test result Y from model (2.1), then drew a TDS dataset of size $n=300$ with an SRC component ($n_0=150$) and a TDC component ($n_1=n_2=n_3=50$) without replacement from the parent cohort. For IPW_{VB}, all variables (Y, X, D) of the patients in the TDS sample as well as Y of all N patients in the parent cohort are used for analysis. For SL_{TDS} and WL_{TDS}, only the variables (Y, X, D) of these patients in the TDS sample ($n=300$) of the same dataset are used for analysis. The results can be found in Table S1 of the supplementary material available at *Biostatistics* online. The IPW_{VB} does not gain much efficiency relative to either SL_{TDS} or WL_{TDS}. In the case of covariate-specific $\mathrm{ROC}_X(t)$, SL_{TDS} is slightly more

Table 2. *Ratio of* $\mathrm{MSE}(SL_{\mathrm{TDS}})$ *under TDS to* $\mathrm{MSE}(PE^{*}_{\mathrm{SRS}})$ *under SRS*

	$t=0.1$	$t=0.3$	$t=0.5$	$t=0.7$	$t=0.9$
$(a_1, a_2) = (\mu_Y - \sigma_Y, \mu_Y + \sigma_Y)$, $(n_0, n_1, n_2, n_3) = (150, 50, 50, 50)$					
$\widehat{\mathrm{ROC}}_{X=0}(t)$	0.65	0.72	0.35	0.18	0.06
$\widehat{\mathrm{ROC}}_{X=1.2}(t)$	0.78	0.43	0.24	0.13	0.04
$\widehat{\mathrm{ROC}}(t)$	0.79	0.82	0.78	0.83	0.80
$(a_1, a_2) = (\mu_Y - \sigma_Y, \mu_Y + \sigma_Y)$, $(n_0, n_1, n_2, n_3) = (150, 75, 0, 75)$					
$\widehat{\mathrm{ROC}}_{X=0}(t)$	0.59	0.64	0.32	0.15	0.06
$\widehat{\mathrm{ROC}}_{X=1.2}(t)$	0.69	0.39	0.21	0.13	0.04
$\widehat{\mathrm{ROC}}(t)$	0.63	0.64	0.65	0.67	0.60
$(a_1, a_2) = (\mu_Y - 1.5\sigma_Y, \mu_Y + 1.5\sigma_Y)$, $(n_0, n_1, n_2, n_3) = (150, 75, 0, 75)$					
$\widehat{\mathrm{ROC}}_{X=0}(t)$	0.56	0.60	0.29	0.15	0.06
$\widehat{\mathrm{ROC}}_{X=1.2}(t)$	0.67	0.37	0.21	0.09	0.04
$\widehat{\mathrm{ROC}}(t)$	0.67	0.68	0.60	0.54	0.46

efficient than IPW_{VB}, likely because SL_{TDS} makes more efficient use of available information. The same conclusion is true when the size of the parent cohort is increased from $N = 1000$ to 5000.

7. Example

As described in Section 1, the investigators of CALGB 30801 are interested in validating the prognostic value of COX2 using data from the patients treated by standard chemotherapy alone. Because the performance of COX2 in the range of moderate and high scores is of primary interest, in order to reduce the number of COX2 negatives and the cost of following them for long-term survival in the validation cohort, the investigators decided to assign standard chemotherapy alone to about one-fourth of all screened COX2 negatives. This design feature makes the percentages of the COX2 negatives, moderates, and positives in the validation cohort disproportional to those of these subgroups in the target population. The investigators plan to continuously enroll negatives to standard chemotherapy until 54 positives (1/4) of the total 216 positives have been randomized. The number of negatives receiving standard chemotherapy alone is estimated approximately 120 and these patients will be followed for long-term survival while the rest of 360 negatives will be off study. The following three types of patients form the SRC: all negatives when one-fourth of the positives are enrolled, all moderates when the one-fourth of the positives are enrolled, and the first 54 positives treated by standard chemotherapy alone. The expected sizes of the three SRC strata are $n_{0,1} = 120$, $n_{0,2} = 26$, and $n_{0,3} = 54$, respectively. The rest of patients receiving standard chemotherapy alone, including $n_1 = 0$ negatives, $n_2 = 78$ moderates, and $n_3 = 54$ positives patients, forms the TDC of the TDS design. It should to be noted that the sizes $\{n_{0,1}, n_{0,2}, n_{0,3}, n_1, n_2, n_3\}$ are the expected values and the actual numbers may vary from trial to trial.

The CALGB trial is ongoing and the data are not yet available for analysis. To illustrate the proposed TDS methods, we simulated the trial data with 800 patients pre-registered for COX2 scoring. The parameters for covariates distribution and predictors associated with survival time were estimated from the data of a preliminary COX2 study (Edelman *and others*, 2007). We generated COX2 from $Beta$ distribution with $\alpha = 2.75$ and $\beta = 1.0$ and scaled its range from [0, 1] to [0, 10]. We generated Age from normal distribution $N(65, 5^2)$, and Male from $Bernoulli(0.65)$. The survival time T was generated from an exponential survival model with regression coefficients for Age, Male, Age × COX2 and Male × COX2 of log(1.3), log(1.5), log(1.2), and log(2), and Age and COX2 were both subtracted by their means.

Table 3. *ROC curve estimates and 95% confidence intervals*

	Method	$t = 0.1$	$t = 0.3$	$t = 0.5$	$t = 0.7$
(Male, Age)			$\hat{\text{ROC}}_X(t)$		
(1, 70)	SL_{TDS}	0.759 [0.680, 0.824]	0.916 [0.876, 0.944]	0.968 [0.949, 0.98]	0.990 [0.983, 0.994]
	WL_{TDS}	0.773 [0.627, 0.874]	0.913 [0.802, 0.965]	0.963 [0.872, 0.990]	0.986 [0.902, 0.998]
	IPW_{VB}	0.804 [0.659, 0.897]	0.933 [0.834, 0.975]	0.974 [0.902, 0.993]	0.991 [0.935, 0.999]
(1, 65)	SL_{TDS}	0.554 [0.426, 0.676]	0.792 [0.682, 0.872]	0.900 [0.826, 0.945]	0.960 [0.922, 0.980]
	WL_{TDS}	0.566 [0.435, 0.735]	0.782 [0.674 , 0.898]	0.885 [0.799, 0.960]	0.948 [0.880, 0.988]
(0, 65)	SL_{TDS}	0.336 [0.197, 0.510]	0.600 [0.417, 0.758]	0.765 [0.596, 0.877]	0.883 [0.761, 0.947]
	WL_{TDS}	0.323 [0.190, 0.493]	0.561 [0.399, 0.711]	0.718 [0.554, 0.840]	0.842 [0.684, 0.929]
			$\hat{\text{ROC}}(t)$		
	SL_{TDS}	0.403 [0.315, 0.498]	0.714 [0.611, 0.798]	0.872 [0.787, 0.926]	0.957 [0.904, 0.981]
	WL_{TDS}	0.426 [0.305, 0.547]	0.699 [0.590, 0.808]	0.845 [0.758, 0.932]	0.935 [0.875, 0.994]
	NP_{TDS}	0.418 [0.269, 0.567]	0.713 [0.556, 0.870]	0.877 [0.756, 0.998]	0.891 [0.774, 1.00]

In analyzing the hypothetical dataset, we formulated the binormal ROC model as in (2.1) of the main article by letting $Y = \text{COX2}$, $D = I(T < 6)$ an indicator for death within 6 years, $X_1 = \text{Age}$ and $X_2 = 1$ for male, 0 for female. All patients will be followed for survival for more than 6 years and we expect that few patients will be censored before 6 years. We also allowed a pairwise interaction between D and (X_1, X_2). To examine the appropriateness of the ROC binormal model, we plotted the residual distribution and found that it followed approximately a normal distribution conditional on D using $Y^{2/3}$ rather than Y. Table 3 shows the estimates and the 95% confidence intervals of ROC curve by the empirical likelihood method SL_{TDS}, the weighted likelihood method WL_{TDS} and the non-parametric method NP_{TDS}, where the confidence intervals for WL_{TDS} and NP_{TDS} were obtained by the bootstrap method. Treating the SRC subjects and the TDC subjects of each COX2 region as separate stratum, we sampled with replacement of sizes n_0, n_1, n_2, n_3, respectively, from these strata of the original sample. Standard errors of the sampling distributions of the bootstrap estimates from 1000 repetitions are used to calculate confidence intervals. Assuming that the test results for all screened patients are known, we applied the IWP method IPW_{VB} and its estimates and confidence intervals are similar to those from WL_{TDS} (not shown). Figure 1 illustrates in the left column that the SL_{TDS} estimates of covariate-specific ROC curve at different combinations of Male (1,0) and Age (65, 70), and in the right column shows the covariate-independent ROC curve with its 95% confidence interval. The curves indicate that the COX2 expressions from Male and Older (70 years old) patients have higher accuracy across all thresholds in predicting whether a patient will remain alive beyond 6 years relative to Female and Younger (65 years old) patients.

8. DISCUSSION

The intent of this article is to propose consistent and efficient estimators for covariate-specific ROC curve and covariate-independent ROC curve when the data arise from a TDS design. The proposed empirical likelihood method avoids a parametric formulation of the nuisance covariate distribution and allows fully efficient inference for the parameters of a binormal ROC model. The proposed empirical likelihood estimators for the covariate-specific ROC curve and the covariate-independent ROC curve have good asymptotic properties. Simulation studies demonstrate acceptable finite sample properties for these estimators as well as better efficiency than the weighted likelihood method and the non-parametric method. As seen in the supplementary material available at *Biostatistics* online, the proposed empirical likelihood estimators perform reasonably well when the binormal assumption is mildly violated. If there is clear evidence of

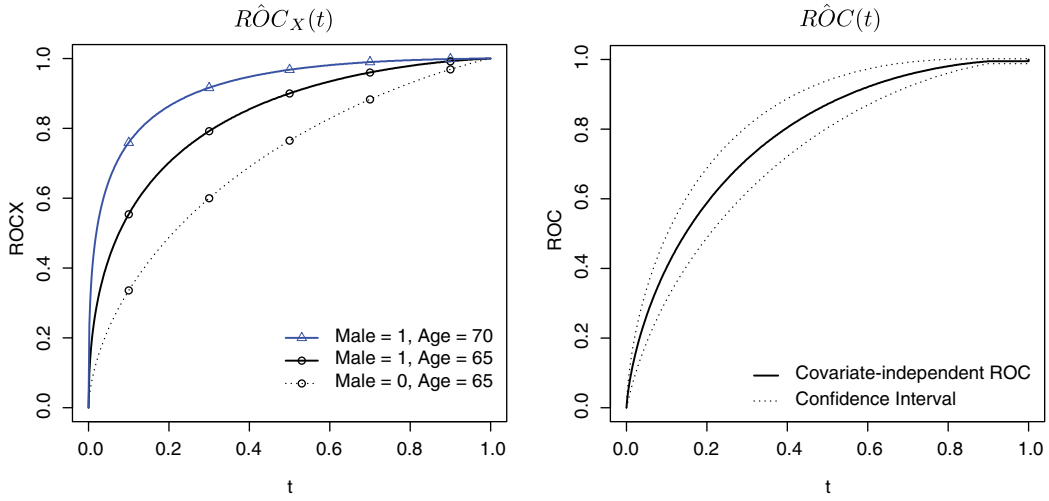

Fig. 1. ROC curve estimated by SL_{TDS} for the COX2 example. Left: $\hat{\text{ROC}}_X(t)$ with $\hat{\text{AUC}}_X = 0.916, 0.83, 0.705$ for $(1, 70), (1, 65), (0, 65)$. Right: $\hat{\text{ROC}}(t)$ with $\hat{\text{AUC}} = 0.744$.

non-normal error, a monotonic increasing transformation can be used to alleviate the problem as the ROC curve is invariant to such transformations.

The TDS design is motivated by the need for improved efficiency for biomarker validation studies with limited resources. The simulation confirms that the TDS design provides more precise estimation of the ROC curve. If the entire ROC curve is of interest, one can gain efficiency by oversampling subjects from the two tails of the distribution of the test result. When the ROC curve at small $t = \text{FPR}$ is of interest, one may oversample subjects in the right tail of the distribution, corresponding to large values of the test result. When the entire ROC curve is of interest, a balanced allocation of subjects to all strata of test results is generally recommended. If the test results for all screened patients are available for analysis, statistical methods for verification bias are applicable. Our simulation shows that the efficiency gain using the IPW method is minimal compared with the proposed TDS methods.

9. Software

The *R* software implementing the proposed methods can be downloaded at http://code.google.com/p/odsroc.

Supplementary material

Supplementary material is available at http://biostatistics.oxfordjournals.org.

Acknowledgment

The authors are grateful to the associate editor and the referee for their insightful comments. *Conflict of Interest:* None declared.

Funding

The research is supported by R03-CA131596, UL1-RR024128, and P01-CA142538.

REFERENCES

ALONZO, T. A. AND PEPE, M. S. (2005). Assessing accuracy of a continuous screening test in the presence of verification bias. *Applied Statistics* **54**, 173–190.

BEGG, C. B. AND GREENES, R. A. (1983). Assessment of diagnostic tests when disease verification is subject to selection bias. *Biometrics* **39**, 207–215.

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.

EDELMAN, M. J., WATSON, D., WANG, X. F., MORRISON, C., KRATZKE, R., JEWELL, S., HODGSON, L., MAUER, A. M., GRAZIANO, S. L., MASTERS, G. A., BEDOR, M., GREEN, M. J. AND VOKES, E. E. (2007). Eicosanoid modulation in advanced lung cancer: COX2 expression is a positive predictive factor for celecoxib + chemotherapy. *Journal of Clinical Oncology* **26**, 848–855.

GRAY, R., BEGG, C. AND GREENES, R. (1984). Construction of receiver operating characteristic curves when disease verification is subject to selection bias. *Medical Decision Making* **4**, 151–164.

LITTLE, R. J. A. AND RUBIN, D. B. (1987). *Statistical Analysis with Missing Data*. New York: Wiley.

METZ, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine* **75**, 237–249.

MORGENTHALER, S. AND VARDI, Y. (1986). Choice-based samples: a nonparametric approach. *Journal of Econometrics* **32**, 109–125.

OWEN, A. B. (2001). *Empirical Likelihood*. New York: Chapman and Hall.

PEPE, M. S. (1997). A regression modelling framework for ROC curves in medical diagnostic testing. *Biometrika* **84**, 595–608.

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

QIN, J. AND LAWLESS, J. F. (1994). Empirical likelihood and general estimating equations. *Annals of Statistics* **22**, 300–325.

QIN, J., ZHANG, B. AND LEUNG, D. H. Y. (2009). Empirical likelihood in missing data problems. *Journal of the American Statistical Association* **104**, 1492–1503.

TOLEDANO, A. L. AND GATSONIS, C. A. (1995). Regression analysis of correlated receiver operating characteristic data. *Academic Radiology* **2**, 30–36.

TOSTESON, A. N. AND BEGG, C. B. (1988). A general regression methodology for ROC curve estimation. *Medical Decision Making* **8**, 204–215.

WANG, X. F., MA, J. L., GEORGE, S. AND ZHOU, H. B. (2012). Estimation of AUC or partial AUC under test-result-dependent sampling. *Statistics in Biopharmaceutical Research* (in press).

WANG, X. F. AND ZHOU, H. B. (2006). A semiparametric empirical likelihood method for biased sampling schemes in epidemiologic studies with auxiliary covariates. *Biometrics* **62**, 1149–1160.

ZHOU, H. B., WEAVER, M. A., QIN, J., LONGNECKER, M. P. AND WANG, M. C. (2002). A semiparametric empirical likelihood method for data from an outcome dependent sampling scheme with a continuous outcome. *Biometrics* **58**, 413–421.

ZHOU, X. H., MCCLISH, D. K. AND OBUCHOWSKI, N. A. (2002). *Statistical Methods in Diagnostic Medicine*. New York: Wiley.

[*Received April 20, 2011; revised May 16, 2012; accepted for publication May 17, 2012*]