Published in final edited form as:
J Am Stat Assoc. 2011 January 1; 106(494): 569–580. doi:10.1198/jasa.2011.tm09807.

Non-parametric Evaluation of Biomarker Accuracy under Nested Case-control Studies

Tianxi Cai
Department of Biostatistics, Harvard University, Boston, MA, USA

Yingye Zheng
Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA, USA

Summary

To evaluate the clinical utility of new risk markers, a crucial step is to measure their predictive accuracy with prospective studies. However, it is often infeasible to obtain marker values for all study participants. The nested case-control (NCC) design is a useful cost-effective strategy for such settings. Under the NCC design, markers are only ascertained for cases and a fraction of controls sampled randomly from the risk sets. The outcome dependent sampling generates a complex data structure and therefore a challenge for analysis. Existing methods for analyzing NCC studies focus primarily on association measures. Here, we propose a class of non-parametric estimators for commonly used accuracy measures. We derived asymptotic expansions for accuracy estimators based on both finite population and Bernoulli sampling and established asymptotic equivalence between the two. Simulation results suggest that the proposed procedures perform well in finite samples. The new procedures were illustrated with data from the Framingham Offspring study.

Keywords

Biomarker study; *Classification Accuracy*; *Conditional Kaplan Meier*; *Inverse Probability Weighting*; *Nested case-control study*; *Predictive Value*; *ROC Curve*; *Time-dependent Accuracy*

1 Introduction

Establishing reliable and parsimonious classification rules for predicting patient survival is a crucial step in the path toward personalized medicine. With the advancement of technology, much progress has been made in identifying new markers useful for disease prognosis. For example, the MammaPrint genetic test holds great potential in predicting disease progression for lymph node negative breast cancer patients and was approved for clinical usage in 2007 (Food and Drug Administration, 2007). In epidemiology studies, risk scores have been developed for many diseases and adopted in public health practice to assist in prevention and treatment efforts. Examples include the Framingham risk score for cardiovascular events (Wilson et al., 1998) and the Gail model for breast cancer (Gail et al., 1989). Here and in the sequel, the terms “biomarker” and “marker” refer generally to the continuous output of a prognostic classifier, such as a biological marker, a genetic score or a clinical risk score.

Several large cohorts have been assembled over the past decade in which biological specimens were collected and stored for future studies. While such prospective cohort studies are crucial for evaluating the prognostic potential of a novel marker, it is often undesirable and/or infeasible to measure markers for the entire cohort due to costs associated with the measurement. Two subcohort sampling designs, the case cohort and the nested

case–control (NCC), are often employed as cost-effective alternatives to the full-cohort design. In particular, under the NCC design, markers are only measured for cases and a fraction of controls selected from the risk sets of the corresponding cases. Such a design is often preferred in a biomarker study as it naturally accommodates practical issues such as batch effects, storage effects and freeze-thaw cycles (Rundle et al., 2005). However, the design also generates complex datasets in which the missingness of the marker values depends on the outcome of interest, making inference about the predictive accuracy of the marker challenging.

Statistical methods for quantifying the prognostic accuracy of a marker with data from NCC studies are not well developed. Existing literature on NCC studies focuses primarily on relative risk parameters. Inference procedures for the hazard ratio under the Cox model (Cox, 1972) have been proposed (Goldstein & Langholz, 1992; Samuelsen, 1997; Chen, 2001). For example, Goldstein & Langholz (1992) developed a conditional logistic regression estimator and Samuelsen (1997) proposed an inverse probability weighted (IPW) estimator. However, such relative measures ignore some fundamental aspects of risk prediction and do not fully capture the predictiveness of a marker (Pepe et al., 2004; Ware, 2006). To construct more clinically relevant accuracy measures, various time-dependent accuracy measures, including the time specific true positive rate (TPR), false positive rate (FPR), receiver operating characteristic (ROC) curve, positive predictive value (PPV) and negative predictive value (NPV), have been proposed (Heagerty et al., 2000; Heagerty & Zheng, 2005; Cai et al., 2006; Zheng et al., 2008). These measures extend existing classification measures for binary outcomes to incorporate the time domain by dichotomizing the continuous event time T into two disease states at any given time-point of interest t. For example, one may consider the classification between subjects with $T \leq t$ and those with $T > t$. This leads to the following accuracy measures:

$$\mathrm{TPR}_t(c) = P(Y \geq c | T \leq t), \quad \mathrm{FPR}_t(c) = P(Y \geq c | T > t),$$

$$\mathrm{PPV}_t(c) = P(T \leq t | Y \geq c), \quad \mathrm{NPV}_t(c) = P(T > t | Y < c),$$

using the convention that a larger value of Y is associated with higher risk of failure. The corresponding time-dependent ROC curve is then $\mathrm{ROC}_t(u) = \mathrm{TPR}_t\left\{\mathrm{FPR}_t^{-1}(u)\right\}$. Existing estimators for these accuracy measures are limited to the case when Y is fully observable. Because of the non-random missingness in Y, they are not directly applicable to data from NCC studies.

Here, we propose non-parametric IPW estimators for the aforementioned accuracy measures with observations inversely weighted by their probabilities of being sampled into the NCC subcohort. We take a non-parametric approach here for the following reason: although risk scores used in practice are often derived from regression models such as the Cox model, validating their prediction performance ideally should not require stringent model assumptions. Our approach is robust in that it remains valid even when the regression model from which the score is derived fails to hold. We consider two different sampling schemes for selecting controls from the risk sets based on: (i) finite population sampling ($\mathbb{F}$-sampling); and (ii) independent Bernoulli sampling ($\mathbb{B}$-sampling). For (i), we obtain IPW estimators using true sampling weights and calculated their asymptotic variance by accounting for the between subject correlation due to sampling. For (ii), we construct IPW estimators using estimated sampling weights. We show that these two types of estimators are equivalent with respect to their asymptotic variance. Such an equivalence has been

established in Breslow & Wellner (2006) for IPW estimators under two-phase stratified case-cohort sampling where the number of matched case-control strata is finite. Here, under NCC design, the number of case-control strata increases with sample size and we show that such an equivalence remains. In addition we consider estimators that accommodate different censoring assumptions. In practice, censoring time C can be quite frequently dependent on marker values. For example, subjects with lower marker values might drop out of the study earlier. To incorporate marker-dependent censoring, we propose IPW kernel smoothing based estimators under the standard survival analysis assumption that T and C are independent given Y. When C is independent of both T and Y, we derive a double IPW estimator as a simple alternative.

The rest of the paper is organized as follows. Section 2 discusses estimation procedures under both sampling schemes and under two types of censoring assumptions. Detailed inference procedures are provided in section 3. We present simulation results in section 4 to demonstrate the finite sample performance of proposed procedures. These procedures are applied to data from the Framingham Offspring study to evaluate the accuracy of a recently develop risk score for predicting cardiovascular events. Concluding remarks are presented in section 5.

2 Estimation

2.1 Sampling Probabilities for the NCC subcohort

Suppose we have a cohort of n individuals followed prospectively for a clinical event. Due to censoring, for T, we observe a bivariate vector (X, δ), where $X = T \wedge C$, $\delta = I(T \le C)$. Let $\mathscr{D} = \{(X_i, \delta_i, Y_i), i=1, \cdots, n\}$ denote the full cohort data, where Y_i only observable if subject i was selected into the NCC subcohort. We assume that Y has a finite support $[\varsigma_l, \varsigma_r]$, C has a finite support $[0, \tau]$. We consider the prediction of survival up to $\tau_0 < \tau$ such that $\inf_{y \in [\varsigma_l, \varsigma_r]} P(X > \tau_0 | Y = y) > 0$. Throughout, we require the standard conditional independent censoring assumption, i.e. T and C are independent given Y. Note that in a purely non-parametric setting, the distribution of T is not identifiable if C is dependent on T given Y and this assumption is not verifiable in general without additional assumptions on the dependence structure (Tsiatis, 1975).

Without loss of generality, we consider a typical NCC study where all cases are included in the subcohort. For each observed case failed at t_j, m controls are randomly sampled from his/her risk set excluding the candidate case, which is of size $\mathfrak{n}^-(t_j) = \sum_{i=1}^{n} I(X_i \ge t_j) - 1$. The m controls are sampled *without* replacement for $\mathbb{F}$-sampling. For $\mathbb{B}$-sampling, each eligible subject in the risk set of t_j is sampled independently with probability $m/\mathfrak{n}^-(t_j)$ as a control for t_j. Both sampling schemes are easy to implement in practice, but $\mathbb{F}$-sampling may be more frequently used. For either of the sampling scheme, V_{0i} denotes whether subject i is ever sampled as a control and $V_i = \delta_i + (1-\delta_i) V_{0i}$ indicates being sampled into the NCC subcohort.

Under $\mathbb{F}$-sampling, the sampling probability for subject i is $\tilde{p}_i = P(V_i = 1 | \mathscr{D}) = \delta_i + (1 - \delta_i) \tilde{p}_{0i}$ (Samuelsen, 1997), and thus the weight used for the IPW estimators is

$$\widehat{w}_i^{\mathbb{F}} = V_i / \tilde{p}_i = \delta_i + (1 - \delta_i) V_{0i} / \tilde{p}_{0i}.$$

where $\tilde{p}_{0i} = 1 - \tilde{G}_m(X_i)$ is the probability of subject i being sampled as a control and

$$\tilde{G}_m(t) = \prod_{j:X_j \le t, \delta_j = 1} \left\{ 1 - \frac{m\delta_j}{\mathfrak{n}^-(X_j)} \right\},$$

For the $\mathbb{B}$-sampling scheme, Let B_{jk} denote an indicator that takes the value 1 if subject k is sampled as a control for subject j (0 otherwise). Then $\{B_{jk}\}$ are independent Bernoulli random variables with success probability $I(X_k \ge X_j, k \ne j)\delta_j m/\mathfrak{n}^-(X_j)$, and $V_{0i} = 1 - \Pi_{j:X_j \le X_i, \delta_j = 1}(1 - B_{ij})$. Note that the true sampling probability for subject i is also $\tilde{p}_i$, i.e., $P(V_i = 1 | \mathscr{D}) = \tilde{p}_i$, and one may use the true sampling weight, i.e., $V_i/\tilde{p}_i$ to construct IPW estimators. However, similar to the findings for case-cohort studies (Breslow & Wellner, 2006; Nan et al., 2009), it can be shown that using *estimated* sampling weights yields improved efficiency. To construct IPW estimators under $\mathbb{B}$-sampling that correspond to those obtained under the $\mathbb{F}$-sampling, we instead use weights

$$\widehat{w}_i^{\mathbb{B}} = V_i/\widehat{p}_i = \delta_i + (1 - \delta_i) V_{0i}/\widehat{p}_{0i},$$

where we estimate $P(V_i = 1 | \mathscr{D})$ as $\widehat{p}_i = \delta_i + (1 - \delta_i)\widehat{p}_{0i}$, $\widehat{p}_{0i} = 1 - \widehat{G}_m(X_i)$ and

$$\widehat{G}_m(t) = \prod_{j:X_j \le t, \delta_j = 1} \left\{ 1 - \frac{\sum_{l=1}^{N} B_{jl}}{\mathfrak{n}^-(X_j)} \right\}$$

2.2 IPW Conditional Nelson-Aalen Estimators of the Conditional Risk and Accuracy Functions

Estimators of the accuracy measures can be constructed by consistently estimating the bivariate survival function $\mathscr{S}(c, t) = P(T > t, Y > c)$ and the marginal distribution of Y, $\mathscr{F}(y) = P(Y \ge y)$. We first direct attention to estimating the conditional survival $S_Y(t) = P(T \ge t \mid Y = y)$. In the following, the IPW weight to account for sampling will be chosen as $\widehat{w}_i = \widehat{w}_i^{\mathbb{B}}$ for $\mathbb{B}$-sampling and $\widehat{w}_i = \widehat{w}_i^{\mathbb{F}}$ for $\mathbb{F}$-sampling.

Conditional Survival Estimation—To estimate $S_Y(t)$ without imposing any parametric assumptions on the relationship between T and Y, we consider the kernel-smoothed conditional Nelson-Aalen (CNA) estimator (Beran, 1981; Dabrowska, 1989; Du & Akritas, 2002). Aside from providing a natural estimator for estimating $S_Y(t)$ nonparametrically, the CNA estimator is also known for its robustness when censoring is dependent on Y. Under NCC sampling, we propose to modify the estimator with IPW to account for the outcome-dependent missingness in Y. Specifically, we propose to estimate the cumulative hazard function $\Lambda_Y(t) = -\log S_Y(t)$ as

$$\widehat{\Lambda}_y(t) = \int_0^t \frac{d\widehat{N}_y(u)}{\widehat{\pi}_y(u)},$$

where $\widehat{N}_y(t) = n^{-1}\sum_{i=1}^{n} \widehat{w}_i K_h(Y_i - y) I(X_i \le t)\delta_i$, $\widehat{\pi}_y(t) = n^{-1}\sum_{i=1}^{n} \widehat{w}_i K_h(Y_i - y) I(X_i \ge t)$, $K_h(x) = K(x/h)/h$ and K is a known smooth symmetric density function. As for the standard

kernel estimation (e.g. Beran, 1981; Dabrowska, 1989; Du & Akritas, 2002), the bandwidth parameter h is assumed to be of order $O(n^{-\nu})$ with $\nu \in [1/5, 1/2)$ to ensure the consistency and asymptotic normality of $\widehat{\Lambda}_y(t)$. More discussions on the order of h for the accuracy estimators are given in section 3. Subsequently, $S_y(t)$ can be estimated as $\widehat{S}_y(t) = \exp\left\{-\widehat{\Lambda}_y(t)\right\}$.

Accuracy Measure Estimation—Based on the estimated conditional survival, we construct the following empirical estimator for the bivariate survival function $\mathscr{S}(c,t) = P(T \geq t, Y \geq c)$ as

$$\widehat{\mathscr{S}}(c,t) = \int_c^{\infty} \widehat{S}_y(t)\, d\widehat{\mathscr{F}}(y) = \frac{\sum_{i=1}^{n} \widehat{S}_{Y_i}(t)\, \widehat{w}_i I(Y_i \geq c)}{\sum_{i=1}^{n} \widehat{w}_i} \tag{2.1}$$

for $\mathscr{S}(t,c)$, where we estimate the marginal distribution of Y as

$$\widehat{\mathscr{F}}(y) = \sum_{i=1}^{n} \widehat{w}_i I(Y_i \geq y) / \sum_{i=1}^{n} \widehat{w}_i. \tag{2.2}$$

Subsequently, we may estimate the marginal survival distribution of T, $\mathscr{S}(t) = P(T \geq t)$, as $\widehat{\mathscr{S}}(t) = \widehat{\mathscr{S}}(c_l, t)$. With the joint and marginal distributions of Y and T estimated, we may easily construct the following plug-in estimators of the aforementioned accuracy measures:

$$\widehat{\mathrm{TPR}}_t(c) = \frac{\left\{1 - \widehat{\mathscr{F}}(c)\right\} - \widehat{\mathscr{S}}(c,t)}{1 - \widehat{\mathscr{S}}(t)}, \qquad \widehat{\mathrm{FPR}}_t(c) = \frac{\widehat{\mathscr{S}}(c,t)}{\widehat{\mathscr{S}}(t)}, \tag{2.3}$$

$$\widehat{\mathrm{PPV}}_t(c) = \frac{\left\{1 - \widehat{\mathscr{F}}(c)\right\} - \widehat{\mathscr{S}}(c,t)}{1 - \widehat{\mathscr{F}}(c)}, \qquad \widehat{\mathrm{NPV}}_t(c) = \frac{\widehat{\mathscr{S}}(t) - \widehat{\mathscr{S}}(c,t)}{\widehat{\mathscr{F}}(c)}, \tag{2.4}$$

The ROC curve can be estimated as $\widehat{\mathrm{ROC}}_t(u) = \widehat{\mathrm{TPR}}_t\left\{\widehat{\mathrm{FPR}}_t^{-1}(u)\right\}$.

2.3 Double Inverse Probability Weighted Estimators

When the censoring C is independent of both Y and T, one may consistently estimate the accuracy measures using double IPW (DIPW) to account for missingness due to both NCC sampling and censoring. Specifically, let $\widehat{\omega}_i^{\dagger}(t) = \delta_i I(X_i \leq t) / \widehat{\mathscr{G}}(X_i) + I(X_i > t) / \widehat{\mathscr{G}}(t)$, where $\widehat{\mathscr{G}}(t)$ is the Kaplan-Merier estimator of $\mathscr{G}(t) = P(C \geq t) = P(C \geq t | Y)$. It is straightforward to see that $E\{\delta_i I(X_i \leq t) / \mathscr{G}(X_i) + I(X_i > t) / \mathscr{G}(t) | T_i, Y_i\} = 1$ and one may use $\widehat{\omega}_i^{\dagger}(t)$ to account for missing information about $I(T_i \leq t)$ due to censoring. Thus, $\mathscr{S}(c,t) = P(T \geq t, Y \geq c)$ and $F(c)$ can be estimated as

$$\widehat{\mathscr{S}}^{\dagger}(t,c) = \frac{\sum_{i=1}^{n} \widehat{\omega}_i \widehat{\omega}_i^{\dagger} I(X_i \geq t, Y_i \geq c)}{\sum_{i=1}^{n} \widehat{\omega}_i \widehat{\omega}_i^{\dagger}}, \qquad \widehat{F}^{\dagger}(c) = \frac{\sum_{i=1}^{n} \widehat{\omega}_i \widehat{\omega}_i^{\dagger} I(Y_i < c)}{\sum_{i=1}^{n} \widehat{\omega}_i \widehat{\omega}_i^{\dagger}},$$

respectively. Subsequently, we obtain estimates of the accuracy measures by replacing $\widehat{\mathscr{S}}(t,c)$, $\widehat{\mathscr{S}}(t)$, $\widehat{F}(c)$ in (2.3) and (2.4) with $\widehat{\mathscr{S}}^{\dagger}(t,c)$, $\widehat{\mathscr{S}}^{\dagger}(t)$, and $\widehat{F}^{\dagger}(c)$, respectively. Note that double weighting was used for $\widehat{F}^{\dagger}(c)$ to ensure that the estimated accuracy measures are between 0 and 1. The resulting estimators are denoted by $\widehat{\mathrm{TPR}}_t^{\dagger}(c)$, $\widehat{\mathrm{FPR}}_t^{\dagger}(c)$, $\widehat{\mathrm{PPV}}_t^{\dagger}(c)$, $\widehat{\mathrm{NPV}}_t^{\dagger}(c)$.

The DIPW approach has the advantage of being simple to calculate without kernel smoothing. However, as shown in the simulation section, the resulting estimators are subject to bias when the censoring distribution depends on the marker values.

3 Asymptotic Properties and Inference Procedures

The NCC sampling scheme brings in additional complexity and poses a significant challenge in the theoretical study of the proposed estimators. Specifically, our proposed estimators based on $\mathbb{F}$-sampling involve the sampling variables $\{V_1^{\mathbb{F}}, \ldots, V_n^{\mathbb{F}}\}$, which are weakly dependent conditional on $\mathscr{D}$. To establish the consistency and asymptotic normality of the proposed estimators, one may account for the dependence using the law of large numbers and central limit theorems for sequences of asymptotically linear negative quadrant dependent random variables (Zhang, 2000; Cai, 2005). Under $\mathbb{B}$-sampling, the sampling variables are independent conditional on $\mathscr{D}$ and the asymptotic properties of the corresponding estimators can be established using empirical processes theory.

Variance Form

For the $\mathbb{F}$-sampling scheme, we show in Appendix A.1 that the asymptotic variance (aVAR) of a generic IPW estimator of the form $n^{-\frac{1}{2}}\sum_{i=1}^{n}\widehat{w}_i^{\mathbb{F}}\xi(\mathbf{D}_i)$ is σ_ξ^2, which is defined in (A.1). In addition, we demonstrate in Appendix A.2 that under $\mathbb{B}$-sampling scheme, the aVAR of the IPW estimator $n^{-\frac{1}{2}}\sum_{i=1}^{n}\widehat{w}_i^{\mathbb{B}}\xi(\mathbf{D}_i)$ is also σ_ξ^2. It is easy to show that for $\mathbb{B}$-sampling, the aVAR of $n^{-\frac{1}{2}}\sum_{i=1}^{n}V_i^{\mathbb{B}}/\tilde{p}_i\xi(\mathbf{D}_i)$, the IPW estimator with true weights, is $E\left\{\xi(\mathbf{D}_i)^2/p_i\right\}>\sigma_\xi^2$, where p_i is defined in Lemma 1. Similar phenomenon has been observed for various IPW estimators (e.g. Breslow & Wellner, 2006; Nan et al., 2009) with case-cohort studies. As emphasized in Robins et al. (1994), the variance form of this type of IPW estimators can be viewed as a residual sum of squares and thus the enrichment of the model for the sampling probability is likely to enhance the efficiency of the estimation. Since the use of $\{\widehat{w}_i^{\mathbb{F}}\}$ and $\{\widehat{w}_i^{\mathbb{B}}\}$ yields the same efficiency, we only provide detailed asymptotic derivations for the latter with $\widehat{w}_i=\widehat{w}_i^{\mathbb{B}}$.

Consistency

In Appendix B, we showed that for the IPW conditional Nelson-Aalen estimator,

$$\sup_{y\in[c_l+h,c_r-h]}|\widehat{\Lambda}_y(t)-\Lambda_y(t)|=O_p\left\{(nh)^{-1/2}\log(n)\right\}$$

when $h=O(n^{-\nu})$ with $\nu\in[1/5,1/2)$. Thus $\widehat{S}_y(t)$ is uniformly consistent for $S_y(t)$ in y. Furthermore, $\widehat{\mathscr{F}}(c)$, $\widehat{\mathrm{TPR}}_t(c)$, $\widehat{\mathrm{FPR}}_t(c)$, $\widehat{\mathrm{TPR}}_t(c)$, and $\widehat{\mathrm{TPR}}_t(c)$, are uniformly consistent for $\mathscr{F}(c)$, $\mathrm{TPR}_t(c)$, $\mathrm{FPR}_t(c)$, $\mathrm{NPV}_t(c)$, and $\mathrm{PPV}_t(c)$ for $c\in\Omega_Y=\left[\underline{c},\bar{c}\right]\subset(c_l,c_r)$ and

$t \in \Omega_T = \left[\underline{\tau}, \bar{\tau}\right] \subset (0, \tau_0]$, where $\underline{c}$, $\bar{c}$, $\underline{\tau}$ and $\bar{\tau}$ are constants such that $P\left(Y < \underline{c}\right) P\left(Y > \bar{c}\right) P\left(T < \underline{\tau}\right) P\left(T > \bar{\tau}\right) > 0$.

Asymptotic Normality and Interval Estimation

To construct confidence intervals (CIs) for the proposed accuracy measures, we show in Appendix C that $\widehat{\mathscr{W}}_{\mathscr{S}}(c,t) = n^{\frac{1}{2}}\left\{\widehat{\mathscr{S}}(c,t) - \mathscr{S}(c,t)\right\}$ and $\widehat{\mathscr{W}}_{\mathscr{F}}(c,t) = n^{\frac{1}{2}}\left\{\widehat{\mathscr{F}}(c) - \mathscr{F}(c)\right\}$ converge jointly to zero-mean Gaussian processes in $c \in \Omega_Y$ when $h = O(n^{-\nu})$ with $\nu \in (1/4, 1/2)$. It is important to note that we require the standard under-smoothing assumption to avoid bias for the resulting accuracy estimators as for smoothed empirical processes (van der Vaart, 1994; Zheng et al., 2008; León et al., 2009). Furthermore, we established weak convergence for the accuracy estimators. With the aforementioned rate for h, the asymptotic distribution of the accuracy estimators does not depend on h at the first order.

The aVAR of these estimators can be estimated empirically, and the CIs can be constructed based on normal approximations. For example, we showed in Appendix C that

$$\widehat{\mathscr{W}}_{\mathrm{FPR}_t}(c) = n^{\frac{1}{2}}\left\{\widehat{\mathrm{FPR}}_t(c) - \mathrm{FPR}_t(c)\right\} \to N\left(0, \sigma^2_{\mathrm{FPR}_t}(c)\right)$$

in distribution, where $\sigma^2_{\mathrm{FPR}_t}(c) = E\left\{\zeta_{\mathrm{FPR}_t}(c;\mathbf{D}_i)^2 / p_i\right\} - m\int \pi(u)^{-1} \eta_{\zeta_{\mathrm{FPR}_t}}(c,u)^2 d\Lambda_{\mathrm{NCC}}(u)$, $\zeta_{\mathrm{FPR}_t}(c;\mathbf{D}_i)$ is defined in Appendix C and $\eta_{\zeta \mathrm{FPR}_t}(c, u) = E\{\zeta_{\mathrm{FPR}_t}(c, \mathbf{D}_i) I(X_i \geq u)(1 - p_i)/p_i\}$. A 95% confidence interval for $\mathrm{FPR}_t(c)$ may be obtained as $\widehat{\mathrm{FPR}}_t(c) \pm 1.96 n^{-\frac{1}{2}} \widehat{\sigma}_{\mathrm{FPR}_t}(c)$, where

$$\widehat{\sigma}^2_{\mathrm{FPR}_t}(c) = n^{-1}\sum_{i=1}^{n} \widehat{w}_i \left\{\widehat{\zeta}_{\mathrm{FPR}_t}(c;\mathbf{D}_i)^2 / \widehat{p}_i\right\} - m\int \{\mathfrak{n}^-(u)/n\}^{-1} \widehat{\eta}_{\zeta_{\mathrm{FPR}_t}}(c,u)^2 d\left[-\log\left\{\widehat{G}_m(u)\right\}\right],$$

$\widehat{\eta}_{\zeta_{\mathrm{FPR}_t}}(c,u) = n^{-1}\sum_{i=1}^{n} \widehat{w}_i \widehat{\zeta}_{\mathrm{FPR}_t}(c, \mathbf{D}_i) I(X_i \geq u)(1 - \widehat{p}_i)/\widehat{p}_i\Big\}$, and $\widehat{\zeta}_{\mathrm{FPR}_t}(c, \mathbf{D})$ is obtained by replacing all theoretical quantities in $\zeta_{\mathrm{FPR}_t}(c, \mathbf{D}_i)$ by their empirical counterparts. Similar point-wise CIs can be constructed for the ROC curve as well as the predictive value functions.

Similar arguments could be used to establish the asymptotic properties of the DIPW estimators when C is independent of Y and T. Under this assumption, $\widehat{\mathscr{G}}(t)$ is a uniformly consistent estimator of $\mathscr{G}(t)$, and $n^{\frac{1}{2}}\left\{\widehat{\mathscr{G}}(t) - \mathscr{G}(t)\right\}$ converges weakly to a zero-mean Gaussian process (Kalbfleisch & Prentice, 2002). This, together with similar arguments as given in the Appendices, can be used to establish the consistency and asymptotic normality of the DIPW accuracy estimators.

4 Numerical Studies

4.1 Simulation Studies

Simulation studies were conducted to assess the performance of the proposed inference procedure in finite samples and to compare the accuracy estimators. To this end, we generated Y from a truncated normal such that $Y = \mathrm{sign}\left(\tilde{Y}\right)\min\left(|\tilde{Y}|, 5\right)$ with $\tilde{Y} \sim N(0, 1)$. The

event time T was generated from a Cox model with $\log T = 1 - \log(3) Y/2 + \epsilon$ and ϵ generated from an independent extreme-value distribution. We generated C as $C = \min(C_0, C_1)$ with $C_0 \sim$ Uniform(.5,1). Two configurations were used for C_1 to incorporate (i) independent censoring and (ii) marker dependent censoring. For (i), we let $C_1 \sim 0.1+$ Gamma(2,2); and for (ii), we let $C_1 = e^{Y/10-1} + e^{Y/5}$Gamma(2,5). Both types of censoring lead to about 90% of censoring and event rate of 5% by $t_0 = 0.5$. The cohort sample size was chosen to be 5000, and for each observed case, either 1 or 3 matched controls were selected. For each dataset, we obtained point and interval estimators for the accuracy of Y in predicting the risk of having an event by t_0 based on both the $\mathbb{F}$- and $\mathbb{B}$-sampling. For each configuration, we simulated 1000 datasets to summarize the empirical performance of the proposed estimators.

We first focus on the CNA estimators. In Table 1(a), we present results for FPR, TPR, PPV, and NPV at $c_p = \mathscr{F}^{-1}(p)$ from simulated datasets with 1 matched control and under independent censoring, for $p = 0.2, 0.4, 0.6, 0.8$. First, we note that all estimators have negligible bias; the estimated standard errors (SE) are close to the sampling standard errors, and the 95% CIs have empirical coverage level close to the nominal level. Second, consistent with the theoretical results, the two sampling schemes yield asymptotically equivalent estimators with both the sampling SE and the estimated SE close to each other. In clinical applications, it is often of interest to summarize the overall accuracy of the marker using the area under the ROC curve (AUC) and also to examine the accuracy of a marker with cut-off value selected to achieve a certain level of sensitivity or specificity. In Table 1(b), we present results for AUC as well as for FPR, PPV and NPV at a sensitivity level of 0.90, representing a relatively low level of false negative rate. The proposed point and interval estimates also perform well with respect to bias and coverage levels.

Results for independent censoring with 3 matched controls are shown in Table 2. In addition to observing reasonable performance for propose estimators, we found that an increase in the number of matched controls appears to be most helpful in improving the estimation of the FPR with about 65% of reduction in the variance. The % reduction in the variance is about 30% for the AUC estimation and ranges from 0% to 21% for the TPR estimation. Similar patterns were also observed under the scenario of marker dependent censoring (Table 3).

As discussed in Section 2.3, we may estimate the accuracy measure via the DIPW approach, which has the computational advantage compared to the CNA approach. To compare the performance of these two approaches under both the $\mathbb{F}$- and $\mathbb{B}$-sampling, we generated data from the same models as described above and assessed the percent bias (relative to the truth) and mean squared errors of all the proposed estimators. In Table 4, we summarize the results for the case with $m = 3$. With independent censoring, all the estimators have negligible bias. Gauged by the mean square errors, both the CNA approach and the DIPW approach yield estimators with comparable efficiency. On the other hand, when the censoring distribution depends on the value of Y, the DIPW approach leads to substantially biased estimators with relative bias as high as 15.6%, while the CNA approach always yields consistent estimators with negligible bias.

4.2 Example

The Framingham risk model, based on several clinical factors, is used extensively for detecting risk for coronary heart disease. However it has only moderate levels of sensitivity and specificity. A new risk model, based on both Framingham risk model variables (Wilson et al., 1998) and an inflammation marker, C-reactive protein (CRP), has been developed recently using data from the Women's Health Study (Cook et al., 2006). We illustrate here

how our proposed procedure can be used to evaluate the clinical utility of the cardiovascular risk prediction model using an independent dataset from the Framingham Offspring study (Kannel et al., 1979).

The Framingham Offspring Study was established in 1971 with 5,124 participants who were monitored prospectively on epidemiological and genetic risk factors of CVD. We consider here 1728 female participants who were free of CVD and have CRP measurement and other clinical information at the second examination. The average age of this subset was about 44 years with standard deviation 10. The outcome we considered was the time from exam date to first major CVD event, including CVD-related death. During the follow-up period, 269 participants experienced at least one CVD event and the 5-year event rate was about 2%. Since CRP measurements are complete in the cohort, the Framingham data allows us to illustrate the methods with a real dataset and compare estimators obtained using data from NCC subcohorts to those from the full cohort.

We first calculated the risk score using an algorithm developed previously in Cook et al. (2006), combining information on age, systolic blood pressure, smoking status, high-density lipoprotein (HDL), total cholesterol, medication for hypertension and CRP concentration. The score was derived using a Cox proportional hazards model. To evaluate the clinical utility of the score in a different dataset, it is sensible to seek a procedure that is independent of the original modeling assumption. Our nonparametric procedures fit well for this purpose. To compare different sampling designs, for each design with either 1 or 3 matched controls, we assembled 500 nested case control datasets by repeatedly sampling the matched controls. For each dataset, we obtained the point and interval estimates of accuracy summaries for the new score in predicting the risk of developing CVD events within 5 years since predictor measurements. In Table 5, we report the average of the estimates over the 500 sets from three subsampling settings: (i) the full cohort; (ii) NCC samples with $m = 1$; and (iii) NCC samples with $m = 3$ Since the $\mathbb{B}$-sampling results in asymptotically equivalent estimators, we focus only on the $\mathbb{F}$-sampling. For comparison, results from both the CNA and DIPW method are reported. Since the results are fairly comparable between these methods, below we summarize estimates from the CNA method only.

Across all accuracy measures, the point estimates from all three subsampling settings are close to each other. The sampling variability of these estimators decreases as the number of controls increases. However, similar to the results in simulation studies, the gain in precision is most pronounced in estimates of FPR. It appears that a NCC design with $m = 3$ would yield accuracy estimators with precision comparable to that of the full cohort in most of the cases. The estimated AUC is about 0.75 with standard error about 0.04 and 95% CI (0.67, 0.84) based on NCC samples with $m = 3$. These estimates suggest that the new score incorporating the CRP information has a moderate accuracy in predicting the 5-year risk of CVD events. One utility of the risk score is to recommend preventive strategies such as a statin therapy to patients who are positive on the score-based test. If a low false negative rate, say 10%, is desirable, then a decision rule based on the corresponding threshold would yield about an FPR of 65% (s.e. 14%); PPV of 99% (s.e. 0.3%) and NPV of 2.9% (s.e. 0.8%).

5 Remarks

Ensuring adequate validation of a prediction model is one of the major challenges in prognostic tool development. In this paper, we proposed nonparametric estimators for prognostic accuracy measures of novel markers with data generated by a NCC design within a prospective cohort study. By using a kernel smoothing technique along with IPW, our proposed estimators are robust and broadly applicable to complex settings where censoring

is marker dependent and marker information is missing by design. Results from extensive simulation studies and practical examples suggest that the commonly time-dependent accuracy measures can be estimated well using data from NCC studies. In general, we find that the CNA approach works well for smaller cohort sizes provided that there are a sufficient number of cases. For example, we also conducted simulation studies with n = 2000 using similar setting as those described above but with slightly higher event rate yielding about 300 cases by the end of the study. As shown in Table 6, the proposed point and interval estimates for the accuracy measures perform well under this setting.

Biological samples collected from cohort members in large studies are often limited and should be used as efficiently as possible. Our proposed approach will enable researchers to efficiently utilize existing resources collected in large cohort studies such as the Nurses' Health Study (Colditz et al., 1997) or the Health Professional Follow-up Study (Hunter et al., 1992), while maintaining scientific rigor in validating novel prediction models for patients' future risk and prognosis. Depending on the quantity of interest, it is possible that a 1:1 matching with $m = 1$ provides sufficient estimation precision. The majority of the precision gain due to a larger m contributes to the FPR estimation. When the desired FPR level is low, the width of the CI could be rather small in general and thus one may achieve a reasonable precision for the estimation of most accuracy measures with a small m. In practice, it appears that when $m = 3$, most of the accuracy estimates achieve reasonable efficiencies relative to those obtained from the full cohort. This echos the finding in the literature that for testing the significance of a single binary covariate, the efficiency of a design with m matched controls per case relative to use of all controls is $m/(m + 1)$ (Ury, 1975; Breslow et al., 1983).

We established the asymptotic equivalence between estimators derived under the $\mathbb{F}$-sampling and $\mathbb{B}$-sampling. This suggests that in practice, sampling with and without replacement can lead to estimators with similar efficiency when appropriate weights are used. While we show that asymptotically the variances of the CNA estimators are not influenced by the choice of bandwidth h provided that it has the correct order, in practice the selection of h in a particular dataset requires special attention to ensure stable estimation. When C is independent of Y and T, the DIPW estimator may be a useful alternative to the CNA estimator with advantage of not requiring smoothing and thus may be more stable when the number of cases is not large. However, one needs to use this estimator with caution as they are prone to bias when the censoring pattern changes with the marker values. The current development considers the predictive accuracy for $I(T \leq t)$ at a pre-specified time point t. When there are multiple time points of interest, one may obtain accuracy estimates across all the points. The asymptotic derivations given in the appendix can be used to justify that properly standardized accuracy estimates over time converge jointly to a multivariate normal. This would allow one to construct simultaneous CIs for these parameters to account for multiple comparisons.

Compared with a case-cohort design, individually matched NCC design is known for its weakness that biomarker information on controls is limited to testing the specific study hypotheses. The proposed IPW approach to analyzing NCC data overcomes such design limitations. Indeed, when selected individuals are weighted inversely by their sampling probability, they provide representative data on the entire cohort and can be used for additional evaluation with a different outcome. The IPW approach, however, may not be most efficient. When auxiliary variables are available, it would be interesting to improve the estimation efficiency via augmentation. For example, one may consider efficient estimators along the lines of Robins et al. (1994) or constructing an optimal augmentation procedure within a pre-specified class of functionals as in Bang & Tsiatis (2000) and Bang & Tsiatis

(2002). The work presented here is an initial step toward future development along that direction.

Appendix

Throughout, let $N_i(t) = I(X_i \le t)\delta_i$, $\bar{N}(t) = n^{-1}\sum_{i=1}^{n} N_i(t)$, $\pi(t) = P(X_i \ge t)$,

$$\Lambda_{\mathrm{NCC}}(t) = \int_0^t \frac{dE\{N_i(u)\}}{\pi(u)}, \qquad \text{and} \qquad G_m(t) = \exp\{-m\Lambda_{\mathrm{NCC}}(t)\}$$

We assume that C has a finite support $[0, \tau]$, which is shorter than that of T. The marker Y is assumed to be continuous and bounded. Throughout, unless noted otherwise, the sup over time t is taken over $[0, \tau]$. We use the notation $\lesssim$ to denote bounded up to a constant and $\simeq$ to denote equal up to $o_p(1)$ in the uniform sense unless specified otherwise. For the kernel function K and marker Y, we make the same assumptions as in Du & Akritas (2002), including: (i) K is a symmetric probability density function with finite support and bounded second derivative; (ii) the distribution function of Y, $\mathscr{F}(y) = P(Y \le y)$ has bounded second and third derivatives with $\inf_x f(x) > 0$, where $f(x) = d\mathscr{F}(x)/dx$.

A Equivalence Between the Finite Population Sampling with True Weights and the Bernoulli Sampling with Estimated Weights

Here, we demonstrate that in general, the IPW estimators obtained based on the two sampling schemes are asymptotically equivalent at the first order.

A.1 Asymptotic Variance with Finite Population Sampling

Our proposed estimators based on the $\mathbb{F}$-sampling involve the sampling variables $\{V_1^{\mathbb{F}}, \ldots, V_n^{\mathbb{F}}\}$ which are weakly dependent conditional on $\mathscr{D}$. To establish the consistency and asymptotic normality of the proposed estimators, one may account for the weak dependence using the law of large numbers (Cai, 2005) and central limit theorem theorems (Zhang, 2000) for sequences of asymptotically linear negative quadrant dependent random variables. Here, we focus primarily on the derivation of the asymptotic variances and outline the justification for the following Lemma:

Lemma 1 Let $\xi(\cdot)$ be a given function of $\mathbf{D} = (\mathrm{X}, \delta, \mathrm{Y})^{\mathrm{T}}$ such that $\mathrm{E}\{\xi(\mathbf{D})\} = 0$, $\mathrm{E}\{\xi(\mathbf{D})^2\} < \infty$ and the total variation of $\xi(\mathbf{D})$ is bounded by a constant. Then the random variable $\widehat{\mathscr{L}_\xi^{\mathbb{F}}}$ of the form

$$n^{-\frac{1}{2}}\sum_{i=1}^{n} \widehat{w}_i^{\mathbb{F}} \xi(\mathbf{D}_i) = n^{-\frac{1}{2}}\sum_{i=1}^{n} \left(\widehat{w}_i^{\mathbb{F}} - 1\right)\xi(\mathbf{D}_i) + n^{-\frac{1}{2}}\sum_{i=1}^{n} \xi(\mathbf{D}_i)$$

has asymptotic variance

$$\sigma_\xi^2 = E\left\{\frac{\xi(\mathbf{D}_i)^2}{p_i}\right\} - m\int \frac{\eta_\xi(u)^2 d\Lambda_{NCC}(u)}{\pi(u)} \tag{A.1}$$

where $\eta_\xi(u) = \mathrm{E}\{\xi(\mathbf{D}_i)\mathrm{I}(\mathrm{X}_i > u)(1 - p_i)/p_i\}$ and $p_i = \delta_i + (1 - \delta_i)\{1 - G_m(\mathrm{X}_i)\}$.

To obtain the asymptotic variance, we note that from Samuelsen (1997), $\mathrm{var}\left(\widehat{w}_i^{\mathbb{F}}|\mathscr{D}\right)=(1-\tilde{p}_i/\tilde{p}_i)$, and for $i\neq j$,

$$\begin{aligned}\mathrm{COV}\left(\widehat{w}_i^{\mathbb{F}},\widehat{w}_j^{\mathbb{F}}|\mathscr{D}\right) &= \widehat{\rho}_{ij}\frac{(1-\tilde{p}_{0i})(1-\tilde{p}_{0j})(1-\delta_i)(1-\delta_j)}{\tilde{p}_{0i}\tilde{p}_{0j}}=\widehat{\rho}_{ij}\frac{(1-\tilde{p}_i)(1-\tilde{p}_j)}{\tilde{p}_i\tilde{p}_j}\\ &= -\frac{m}{n}\int\frac{I(X_i\geq t)I(X_j\geq t)(1-p_i)(1-p_j)}{p_ip_j}\frac{d\Lambda_{\mathrm{NCC}}(t)}{\pi(t)}+O_p\left(n^{-3/2}\right).\end{aligned}$$

where

$$\begin{aligned}\widehat{\rho}_{ij} &= \prod_{k:X_k<\min(X_i,X_j)}\left[1-\frac{2m\delta_k}{\mathfrak{n}^-(X_k)}+\frac{m(m-1)\delta_k}{\mathfrak{n}^-(X_k)\{\mathfrak{n}^-(X_k)-1\}}\right]/\left\{1-\frac{m\delta_k}{\mathfrak{n}^-(X_k)}\right\}^2-1\\ &= -\frac{m}{n}\int I(X_i\geq t)\,I\left(X_j\geq t\right)\frac{d\Lambda_{\mathrm{NCC}}(t)}{\pi(t)}+O_p\left(n^{-3/2}\right).\end{aligned}$$

On the other hand, since $E\left\{\widehat{w}_i^{\mathbb{F}}|\mathscr{D}\right\}=1$.

$$\begin{aligned}&\mathrm{var}\left\{n^{-\frac{1}{2}}\sum_{i=1}^n\widehat{w}_i^{\mathbb{F}}\xi(\mathbf{D}_i)\right\}=E\left[\mathrm{var}\left\{n^{-\frac{1}{2}}\sum_{i=1}^n\widehat{w}_i^{\mathbb{F}}\xi(\mathbf{D}_i)|\mathscr{D}\right\}\right]+\mathrm{var}\left[E\left\{n^{-\frac{1}{2}}\sum_{i=1}^n\widehat{w}_i^{\mathbb{F}}\xi(\mathbf{D}_i)|\mathscr{D}\right\}\right]\\ &=E\left[n^{-1}\sum_{i=1}^n\frac{1-\tilde{p}_i}{\tilde{p}_i}\xi(\mathbf{D}_i)^2+n^{-1}\sum_{i\neq j}\widehat{\rho}_{ij}\frac{(1-\tilde{p}_i)(1-\tilde{p}_j)\xi(\mathbf{D}_i)\xi(\mathbf{D}_j)}{\tilde{p}_i\tilde{p}_j}\right]+E\left\{\xi(\mathbf{D}_i)^2\right\}\\ &\simeq E\left\{\frac{\xi(\mathbf{D}_i)^2}{p_i}\right\}-\frac{m}{n^2}\sum_{i\neq j}E\left[\int\frac{I(X_i\geq t)I(X_j>t)(1-p_i)(1-p_j)\xi(\mathbf{D}_i)\xi(\mathbf{D}_j)}{p_ip_j}\frac{d\Lambda_{\mathrm{NCC}(t)}}{\pi(t)}\right]\simeq\sigma_\xi^2\end{aligned}$$

A.2 Bernoulli Sampling with Estimated Weights

Here we derive the asymptotic variance for a statistic of the form

$$\widehat{\mathscr{Z}_\xi^{\mathbb{B}}}=n^{-\frac{1}{2}}\sum_{i=1}^n\widehat{w}_i^{\mathbb{B}}\xi(\mathbf{D}_i)$$

for some deterministic function ξ. First, since conditional on $\mathscr{D}$, $\{B_{ij}\}$ are independent Bernoulli random variables with success probability $I\left(X_j\geq X_i, i\neq j\right)m\delta_i/\mathfrak{n}^-(X_i)$ and $\{V_1^{\mathbb{B}},\ldots,V_n^{\mathbb{B}}\}$ are independent Bernoulli with success probability $\tilde{p}_i$. By the standard empirical process theory (Pollard, 1990), it is not difficult to show that, uniformly over $t\in[0,\tau]$, conditional on $\mathscr{D}$

$$n^{\frac{1}{2}}\frac{\widehat{G}_m(t)-\tilde{G}_m(t)}{\tilde{G}_m(t)}=-n^{\frac{1}{2}}\sum_{i=1}^n\frac{N_i(t)}{\mathfrak{n}^-(X_i)}\left[\sum_{j=1}^n I\left(X_j\geq X_i\right)\left\{B_{ij}-\frac{m}{\mathfrak{n}^-(X_i)}\right\}\right]+O_p\left(n^{-\frac{1}{2}}\right)\quad(A.2)$$

which converges weakly to a zero-mean Gaussian process. This, together with a taylor expansion and a uniform law of large numbers (ULLN) (Pollard, 1990), implies that

$$\widehat{\mathscr{Z}_\xi^{\mathbb{B}}}\simeq n^{-\frac{1}{2}}\sum_{j=1}^n\tilde{w}_j^{\mathbb{B}}\xi\left(\mathbf{D}_j\right)-\widehat{\varepsilon}_\xi\quad(A.3)$$

where $\tilde{w}_j^{\mathbb{B}}=V_j^{\mathbb{B}}/\tilde{p}_j$, $\widehat{\varepsilon}_\xi=n^{-\frac{1}{2}}\sum_{i=1}^n\int\eta_\xi(s)\,d\widehat{\varepsilon}_j(s)$ and

$$\widehat{\varepsilon}_j(s)=\sum_{i=1}^{n}\frac{I\left(X_j\geq X_i\right)}{\mathfrak{n}^-(X_i)/n}\left(B_{ij}-\frac{m}{\mathfrak{n}^-(X_i)}\right)N_i(s).$$

We next approximate $\mathrm{var}\left\{\widehat{\mathscr{Z}}_\xi^{\mathbb{B}}|\mathscr{D}\right\}$. Since $\mathrm{var}\left(B_{ij}|\mathscr{D}\right)=m/\mathfrak{n}^-(X_i)+O_p\left(n^{-2}\right)$ and $\mathrm{var}\left(\tilde{w}_i^{\mathbb{B}}|\mathscr{D}\right)=(1-\tilde{p}_i)/\tilde{p}_i$,

$$\mathrm{var}\left\{\left(\tilde{w}_j^{\mathbb{B}}-1\right)\xi\left(\mathbf{D}_j\right)|\mathscr{D}\right\}=n^{-1}\sum_{i=1}^{n}\frac{1-\tilde{p}_j}{\tilde{p}_j}\xi\left(\mathbf{D}_j\right)^2 \tag{A.4}$$

$$\mathrm{var}\left(\widehat{\varepsilon}_\xi|\mathscr{D}\right)=nm\sum_{i=1}^{n}\frac{\eta_\xi(X_i)^2\delta_i}{\mathfrak{n}^-(X_i)^2}+O_p\left(n^{-2}\right)\to m\int\frac{\eta_\xi(t)^2}{\pi(t)}d\Lambda_{\mathrm{NCC}}(t) \tag{A.5}$$

Conditional on $\mathscr{D}$, $B_{ij'}$ and $V_j^{\mathbb{B}}$ are independent when $j'\neq j$. Thus, the covariance between $n^{-\frac{1}{2}}\sum_{i=1}^{n}\left(\tilde{w}_j^{\mathbb{B}}-1\right)\xi\left(\mathbf{D}_j\right)$ and $\widehat{\varepsilon}_\xi$ given $\mathscr{D}$ is

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\eta_\xi(X_i)\delta_i}{\mathfrak{n}^-(X_i)}I\left(X_j\geq X_i\right)\frac{\mathrm{cov}\left(B_{ij},V_j^{\mathbb{B}}|\mathscr{D}\right)\xi\left(\mathbf{D}_j\right)}{\tilde{p}_j}$$

Since $V_{0j}^{\mathbb{B}}=0$ implies $B_{ij}=0$,

$$\begin{aligned}\mathrm{cov}\left(B_{ij},V_j^{\mathbb{B}}|\mathscr{D}\right)&=\mathrm{cov}\left(1-B_{ij},1-V_j^{\mathbb{B}}|\mathscr{D}\right)=\left(1-\delta_j\right)\left\{P\left(V_{0j}^{\mathbb{B}}=0\right)-P\left(V_{0j}^{\mathbb{B}}=0\right)P\left(B_{ij}=0\right)\right\}\\&=\frac{m(1-\delta_j)P\left(V_{0j}^{\mathbb{B}}=0\right)}{\mathfrak{n}^-(X_i)}=\frac{m\left(1-\tilde{p}_j\right)}{\mathfrak{n}^-(X_i)}\end{aligned}$$

This, together with a ULLN, implies that

$$\mathrm{cov}\left(\tilde{w}_j^{\mathbb{B}}\xi\left(\mathbf{D}_j\right),\widehat{\varepsilon}_\xi|\mathscr{D}\right)\to m\int\frac{\eta_\xi(t)^2d\Lambda_{\mathrm{NCC}}(t)}{\pi(t)}. \tag{A.6}$$

On the other hand, $\mathrm{var}\left[E\left\{\widehat{\mathscr{Z}}_\xi^{\mathbb{B}}|\mathscr{D}\right\}\right]\simeq E\left\{\xi(\mathbf{D}_i)^2\right\}$. Thus, we have

Lemma 2 Let $\xi(\cdot)$ be a given function of $\mathbf{D}$ such that $\mathrm{E}\{\xi(\mathbf{D})\}=0$, $\mathrm{E}\{\xi(\mathbf{D})^2\}<\infty$ and the total variation of $\xi(\mathbf{D})$ is bounded by a constant. Then the random variable $\widehat{\mathscr{Z}}_\xi^{\mathbb{B}}$ of the form

$$n^{-\frac{1}{2}}\sum_{i=1}^{n}\widehat{w}_i^{\mathbb{B}}\xi(\mathbf{D}_i)=n^{-\frac{1}{2}}\sum_{i=1}^{n}\left(\widehat{w}_i^{\mathbb{B}}-1\right)\xi(\mathbf{D}_i)+n^{-\frac{1}{2}}\sum_{i=1}^{n}\xi(\mathbf{D}_i)$$

has asymptotic variance

$$\sigma_\xi^2=E\left\{\frac{\xi(\mathbf{D}_i)^2}{p_i}\right\}-m\int\frac{\eta_\xi(u)^2d\Lambda_{NCC}(u)}{\pi(u)}. \tag{A.7}$$

B Uniform Consistency of the Absolute Risk and Accuracy Estimators under Bernoulli Sampling

For the consistency, we assume that $h = O(n^{-\nu})$ with $\nu \in [1/5, 1/2)$. We first establish the following uniform convergence rate for $\widehat{\Lambda}_y(t_0) = \int_0^{t_0} \{\widehat{\pi}_y(t)\}^{-1} d\widehat{N}_y(t)$:

$$\sup_{y,t_0} |\widehat{\Lambda}_y(t_0 - \Lambda_y(t_0))| = O_p\left\{(nh)^{-\frac{1}{2}} \log(n)\right\}, \tag{B.1}$$

where

$$\widehat{N}_y(t) = \frac{\int K_h(x-y)\, \widehat{H}_N(dx;t)}{\int K_h(x-y)\, \widehat{\mathscr{F}}_0(dx)}, \qquad \widehat{\pi}_y(t) = \frac{\int K_h(x-y)\, \widehat{H}_\pi(dx;t)}{\int K_h(x-y)\, \widehat{\mathscr{F}}_0(dx)},$$

$$\widehat{H}_N(x;t) = n^{-1} \sum_{i=1}^n \widehat{w}_i I(Y_i \le x) N_i(t), \qquad \widehat{H}_\pi(x;t) = n^{-1} \sum_{i=1}^n \widehat{w}_i I(Y_i \le x) I(X_i \ge t)$$

and $\widehat{\mathscr{F}}_0(x) = n^{-1} \sum_{i=1}^n I(Y_i \le x)$. By Lemma A.3 in Bilias et al. (1997), it suffices to show that

$$\sup_{y,t} |\widehat{N}_y(t) - A_y(t)| = O_p\left\{(nh)^{-\frac{1}{2}} \log(n)\right\}, \tag{B.2}$$

$$\sup_{y,t} |\widehat{\pi}_y(t) - \pi_y(t)| = O_p\left\{(nh)^{-\frac{1}{2}} \log(n)\right\}, \tag{B.3}$$

where $A_y(t) = E\{N_i(t) \mid Y_i = y\}$, $\pi_y(t) = P(X_i \ge t \mid Y_i = y)$.

First, we note that since $\widehat{w}_i \delta_i = \delta_i$,

$$\widehat{N}_y(t) - A_y(t) = \frac{\int K_h(x-y)\, \widehat{H}_{0N}(dx;t)}{\int K_h(x-y)\, \widehat{\mathscr{F}}_0(dx)} - A_y(t).$$

where $\widehat{H}_{0N}(x;t) = n^{-1} \sum_{i=1}^n I(Y_i \le x) N_i(t)$. Then (B.2) follows immediately from Du & Akritas (2002). To show (B.3), we let $\widehat{H}_{0\pi}(x;t) = n^{-1} \sum_{i=1}^n I(Y_i \le x, X_i \ge t)$ and write $|\widehat{\pi}_y(t) - \pi_y(t)| \le |\epsilon_{1y}(t)| + |\epsilon_{2y}(t)|$, where

$$\epsilon_{1y}(t) = \frac{\int K_h(x-y) \{\widehat{H}_\pi(dx;t) - \widehat{H}_{0\pi}(dx;t)\}}{\int K_h(x-y)\, \widehat{\mathscr{F}}_0(dx)}$$

and

$$\epsilon_{2y}(t) = \frac{\int K_h(x-y)\,\widehat{H}_{0\pi}(dx;t)}{\int K_h(x-y)\,\widehat{\mathscr{F}}_0(dx)} - \pi_y(t).$$

For $\epsilon_{1y}(t)$, we first note that from a functional central limit theorem (FCLT) (Pollard, 1990), $n^{-\frac{1}{2}}\sum_{i=1}^{n}\widehat{\varepsilon}_j(s)$ converges weakly to a zero-mean Gaussian process in s and thus

$$\sup_{s\le\tau}\left|n^{-\frac{1}{2}}\sum_{i=1}^{n}\widehat{\varepsilon}_j(s)\right| = O_p(1) \tag{B.4}$$

This, together with (A.2), a ULLN and Lemma A.3 of Bilias et al. (1997), yields

$$n^{\frac{1}{2}}\left\{\widehat{H}_\pi(x;t) - \widehat{H}_{0\pi}(x;t)\right\} = n^{-\frac{1}{2}}\sum_{i=1}^{n}\left\{(\tilde{w}_i - 1)\,I(Y_i \le x, X_i \ge t) - \int \eta_{H\pi}(s;x,t)\,d\widehat{\varepsilon}_i(s)\right\} \tag{B.5}$$

where $\eta_{H\pi}(s; x, t) = E\{I(t > X_i \ge s,\ Y_i \le x)(1 - p_i)/p_i\}$. Conditional on $\mathscr{D}$, $\left\{(\tilde{w}_i - 1)\,I(Y_i \le x, X_i \ge t) - \int \eta_{H\pi}(s;x,t)\,d\widehat{\varepsilon}_i(s), i=1,\cdots,n\right\}$ are independent with mean 0. Furthermore, $(\widehat{w}_i - 1)\,I(Y_i \le x, X_i \ge t) - \int \eta_{H\pi}(s;x,t)\,d\widehat{\varepsilon}_i(s)$ can be written as differences of monotone functions with a constant bound and thus has finite psuedo-dimension (Pollard, 1990). Then by a FCLT, conditional on $\mathscr{D}$, $n^{\frac{1}{2}}\left\{\widehat{H}_N(x;t) - \widehat{H}_{0N}(x;t)\right\}$ converges weakly to a zero-mean Gaussian process in (x, t). This, together with the standard arguments given in Bickel & Rosenblatt (1973), implies that $\sup_{t,y}|\epsilon_{1y}(t)| = O_p\left\{(nh)^{-\frac{1}{2}}\log(n)\right\}$. On the other hand, from Du & Akritas (2002), $\sup_{t,y}|\epsilon_{y2}(t)| = O_p\left\{(nh)^{-\frac{1}{2}}\log(n)\right\}$. This concludes the proof for (B.2) and thus we have (B.1).

The convergence of $\widehat{\Lambda}_y(t)$ in (B.1) implies that $\widehat{S}_y(t)$ is uniformly consistent for $S_y(t)$. Since $\widehat{\mathscr{F}}(x) = \widehat{\mathscr{H}}(x;0)$, the uniform consistency of $\widehat{\mathscr{F}}(x)$ for $\mathscr{F}(x)$ follows immediately. The convergences of $\widehat{\mathscr{F}}(x)$ and $\widehat{\mathscr{S}_y}(t)$ along with a continuous mapping theorem and Lemma A.3 of Bilias et al. (1997) imply the uniform consistency of $\widehat{\mathscr{S}}(c, t)$ and all the proposed accuracy measure estimators.

C Asymptotic Distribution of Accuracy Estimators under Bernoulli Sampling

To obtain an asymptotic expansion for the proposed accuracy estimators, we first obtain approximations for $\bar{\mathscr{W}}_y(t) = \widehat{\Lambda}_y(t) - \Lambda_y(t)$ and $\widehat{\mathscr{W}}_{\mathscr{F}}(y) = n^{\frac{1}{2}}\left\{\widehat{\mathscr{F}}_{(y)} - \mathscr{F}(y)\right\}$. To remove the potential bias in the accuracy estimators due to the kernel smoothing, we now require $h = O(n^{-\nu})$ with $1/4 < \nu < 1/2$. From the asymptotic approximations given in Appendix B for $\widehat{N}_y(t)$ and $\widehat{\pi}_y(t)$ as well as the arguments given in Bickel & Rosenblatt (1973), we have

$$\bar{\mathscr{W}}_y(t) = n^{-1}\sum_{i=1}^{n}\left\{\frac{\tilde{w}_i K_h(Y_i - y)\,M_y(t;D_i)}{f(y)} + \int \zeta_\Lambda(t, s; y)\,d\widehat{\varepsilon}_i(s)\right\} + o_p\left(n^{-\frac{1}{2}}\right) \tag{C.1}$$

where $M_y(t;\mathbf{D}_i) = \int_0^t \pi_y(u)^{-1} \left\{ dN_i(u) - I(X_i \geq u)\, d\Lambda_y(u) \right\}$,
$\zeta_\Lambda = (t, s; y) = \int_0^t E\{ I(X_i \geq s \wedge u)(1 - p_i)/p_i | Y_i = y \} \pi_y(u)^{-1} d\Lambda_y(u)$.

Next, noting that $n^{-1}\sum_{i=1}^n \widehat{w}_i \to 1$ in probability, we write

$$\widehat{\mathscr{W}}_{\mathscr{F}}(y) \simeq n^{-\frac{1}{2}} \sum_{i=1}^n (\widehat{w}_i - 1)\{ I(Y_i \leq y) - \mathscr{F}(y) \} + \widehat{\mathscr{W}}_{0\mathscr{F}}(y)$$

where $\widehat{\mathscr{W}}_{0\mathscr{F}}(y) = n^{-\frac{1}{2}} \sum_{i=1}^n \{ I(Y_i \leq y) - \mathscr{F}(y) \}$. From (A.2) and similar arguments as given for the approximation of (A.3), we have
$\widehat{\mathscr{W}}_{\mathscr{F}}(y) \simeq n^{-\frac{1}{2}} \sum_{i=1}^n (\widehat{w}_i - 1)\{ I(Y_i \leq y) - \mathscr{F}(y) \} + \widehat{\mathscr{W}}_{0\mathscr{F}}(y)$, where

$$\widehat{W}_{1\mathscr{F}}(y) = n^{-\frac{1}{2}} \sum_{i=1}^n \left\{ (\tilde{w}_i - 1)\{ I(Y_i \leq y) - \mathscr{F}(y) \} \int \zeta_{\mathscr{F}}(y; s)\, d\widehat{\varepsilon}_i(s) \right\}$$

and $\zeta_{\mathscr{F}}(y; s) = E\{ I(Y_i \leq y, X_i \geq s)(1 - p_i)/p_i \}$. Conditional on $\mathscr{D}$,
$\left\{ (\tilde{w}_i - 1)\{ I(Y_i \leq y) - \mathscr{F}(y) \} - \int \zeta_{\mathscr{F}}(y; s)\, d\widehat{\varepsilon}_i(s), i = 1, \ldots, n \right\}$ are independent with mean 0 and finite pseudo-dimension. Thus, by a FCLT, $\widehat{\mathscr{W}}_{1\mathscr{F}}(\cdot)$ converges weakly to a zero-mean Gaussian process $\mathbb{W}_{1\mathscr{F}}(\cdot)$. On the other hand, $\widehat{W}_{0\mathscr{F}}(y)$ is tight and weakly convergent to a zero-mean Gaussian process $\mathbb{W}_{0\mathscr{F}}(\cdot)$. Since, $n^{-\frac{1}{2}} \sum_{i=1}^n (\widehat{w}_i - 1)\{ I(Y_i \leq y) - \mathscr{F}(y) \}$ and $\widehat{\mathscr{W}}_{0\mathscr{F}}(y)$ are independent conditional on $\mathscr{D}$, $\widehat{\mathscr{W}}_{\mathscr{F}}(y)$ converges weakly to $\mathbb{W}_{1\mathscr{F}}(y) + \mathbb{W}_{0\mathscr{F}}(y)$.

We next approximate the distribution of $\widehat{\mathscr{W}}_s(c, t) = n^{\frac{1}{2}} \left\{ \widehat{S}(c, t) - S(c, t) \right\}$. Since $\widehat{\mathscr{W}}_{\mathscr{F}}(y) = O_p(1)$ and $\bar{\mathscr{W}}_y(t) = O_p\left\{ (nh)^{n^{\frac{1}{2}}} \log(n) \right\}$, we have

$$\widehat{\mathscr{W}}_s(c, t) \simeq -n^{\frac{1}{2}} \int_c^\infty S_y(t)\, \bar{\mathscr{W}}_y(t)\, d\mathscr{F}(y) + \int_c^\infty S_y(t)\, d\widehat{\mathscr{W}}_{\mathscr{F}}(y)$$

It follows from (C.1) that

$$\int_c^\infty S_y(t)\, \bar{\mathscr{W}}_y(t)\, d\mathscr{F}(y) = n^{-1} \sum_{i=1}^n \left\{ \int_c^\infty \frac{\widehat{w}_i K_h(Y_i - y)\, S_y(t)\, M_y(t;\mathbf{D}_i)}{f(y)} dy + \int \zeta_s(c, t, s)\, d\widehat{\varepsilon}_i(s) \right\} + o_p\left(n^{-\frac{1}{2}} \right).$$

where $\zeta_S(c, t, s) = \int_c^\infty \int S_y(t)\, \zeta_\Lambda(t, s; y)\, d\mathscr{F}(y)$. By a change of variable $\psi = (y - Y_i)/h$,

$$n^{-\frac{1}{2}} \sum_{i=1}^n \int_c^\infty \frac{\tilde{w}_i K_h(Y_i - y)\, \mathscr{S}_y(t)\, M_y(t;\mathbf{D}_i)}{f(y)} dy \simeq n^{-\frac{1}{2}} \sum_{i=1}^n \tilde{w}_i I(Y_i > c)\, \mathscr{S}_{Y_i}(t)\, M_{Y_i}(t;\mathbf{D}_i).$$

Therefore,

$$\widehat{\mathscr{W}_S}(c,t) \simeq n^{-\frac{1}{2}} \sum_{i=1}^{n} \left\{ (\widehat{w}_i - 1)\, \mathscr{R}_S(c,t;\mathbf{D}_i) + \int \zeta_S(c,t,s)\, d\widehat{\varepsilon}_i(s) \right\} + n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathscr{R}_S(c,t;\mathbf{D}_i)$$

where $\mathscr{R}_S(c,t;\mathbf{D}_i) = S_{Y_i}(t)\, I(Y_i > c) - S(c,t) - I(Y_i > c)\, S_{Y_i}(t)\, M_{Y_i}(t;\mathbf{D}_i)$. Using similar arguments as given above, a FCLT may be used to show that conditional on $\mathscr{D}$, $n^{-\frac{1}{2}} \sum_{i=1}^{n} (\tilde{w}_i - 1)\, \mathscr{R}_S(c,t;\mathbf{D}_i) + \int \zeta_S(c,t,s)\, d\widehat{\varepsilon}_i(s)$ converges weakly to a zero-mean Gaussian process $\mathbb{W}_{1S}(c,t)$. On the other hand, $n^{-\frac{1}{2}} \sum_{i=1}^{n} \mathscr{R}_S(c,t;\mathbf{D}_i)$ also converges weakly to a zero-mean Gaussian process $\mathbb{W}_{0S}(c,t)$. Therefore, $\widehat{\mathscr{W}_S}(c,t)$ converges weakly to $\mathbb{W}_{1S}(c,t) + \mathbb{W}_{0S}(c,t)$.

Furthermore, it is not difficult to show that the weak convergence of $\widehat{\mathscr{W}_S}(c,t) \to \mathbb{W}_{1S}(c,t) + \mathbb{W}_{0S}(c,t)$ and $\widehat{\mathscr{W}_{\mathscr{F}}}(c) \to \mathbb{W}_{1\mathscr{F}}(c) + \mathbb{W}_{0\mathscr{F}}(c)$ holds jointly. The asymptotic distribution of the accuracy estimators follows directly from the joint distribution of $\widehat{\mathscr{W}_S}(c,t)$ and $\widehat{\mathscr{W}_{\mathscr{F}}}(c)$. This, together with a functional delta theorem, implies the following approximations for $\widehat{\mathscr{W}_{\mathrm{FPR}_t}}(c) = n^{\frac{1}{2}} \left\{ \widehat{\mathrm{FPR}}_t(c) - \mathrm{FPR}_t(c) \right\}$, $\widehat{\mathscr{W}_{\mathrm{TPR}_t}}(c) = n^{\frac{1}{2}} \left\{ \widehat{\mathrm{TPR}}_t(c) - \mathrm{TPR}_t(c) \right\}$, $\widehat{\mathscr{W}_{\mathrm{NPV}_t}}(c) = n^{\frac{1}{2}} \left\{ \widehat{\mathrm{NPV}}_t(c) - \mathrm{NPV}_t(c) \right\}$, and $\widehat{\mathscr{W}_{\mathrm{PPV}_t}}(c) = n^{\frac{1}{2}} \left\{ \widehat{\mathrm{PPV}}_t(c) - \mathrm{PPV}_t(c) \right\}$,

$$\widehat{\mathscr{W}_{\mathrm{FPR}_t}}(c) \simeq \frac{\widehat{\mathscr{W}_S}(c,t) - \mathrm{FPR}_t(c)\, \widehat{\mathscr{W}_S}(\mathfrak{c}_l,t)}{\mathscr{S}(t)} \qquad \widehat{\mathscr{W}_{\mathrm{TPR}_t}}(c) \simeq \frac{\mathrm{TPR}_t(c)\, \widehat{\mathscr{W}_S}(\mathfrak{c}_l,t) - \widehat{\mathscr{W}_{\mathscr{F}}}(c) - \widehat{\mathscr{W}_S}(c,t)}{1 - \mathscr{S}(t)},$$

$$\widehat{\mathscr{W}_{\mathrm{PPV}_t}}(c) \simeq \frac{\{\mathrm{PPV}_t(c) - 1\}\, \widehat{\mathscr{W}_{\mathscr{F}}}(c) - \widehat{\mathscr{W}_S}(c,t)}{1 - \mathscr{F}(c)}, \widehat{\mathscr{W}_{\mathrm{NPV}t}}(c) \simeq \frac{\widehat{\mathscr{W}_S}(t) - \widehat{\mathscr{W}_S}(c,t) - \mathrm{NPV}_t(c)\, \widehat{\mathscr{W}_{\mathscr{F}}}(c)}{\mathscr{F}(c)}.$$

The same arguments as given above can then be used to establish the weak convergence for these processes and obtain the asymptotic variance based on (A.7). For example, since $\widehat{\mathscr{W}_{\mathrm{FPR}_t}}(c) \simeq n^{-\frac{1}{2}} \sum_{i=1}^{n} \widehat{\omega}_i \zeta_{\mathrm{FPR}_t}(c;\mathbf{D}_i)$ with $\zeta_{\mathrm{FPR}_t}(c;\mathbf{D}_i) = S(t)^{-1} \{ \mathscr{R}_S(c,t;\mathbf{D}_i) - \mathrm{FPR}_t(c)\, \mathscr{R}_S(\mathfrak{c}_l,t;\mathbf{D}_i) \}$, $\widehat{\mathscr{W}_{\mathrm{FPR}_t}}(c) \to N\left(0, \sigma^2_{\mathrm{FPR}_t}(c)\right)$ in distribution, where $\sigma^2_{\mathrm{FPR}_t}(c) = E\left\{ \zeta_{\mathrm{FPR}_t}(c;\mathbf{D}_i)^2 / p_i \right\} - m \int \pi(u)^{-1} \eta_{\zeta_{\mathrm{FPR}_t}}(c,u)^2 d\Lambda_{\mathrm{NCC}}(u)$, and $\eta_{\zeta_{\mathrm{FPR}_t}}(c,u) = E\left\{ \zeta_{\mathrm{FPR}_t}(c;\mathbf{D}_i)\, I(X_i \geq u)\, (1 - p_i) / p_i \right\}$.

To establish the weak convergence of the ROC curve estimator, we first note that the arguments above can be extended to show that the weak convergences of the two processes, $\widehat{\mathscr{W}_{\mathrm{FPR}_t}}(c)$ and $\widehat{\mathscr{W}_{\mathrm{TPR}_t}}(c)$, hold jointly. This, together with the stochastic equicontinuity of these processes, implies that for $u \in [u_l, u_r] \subset (0,1)$,

$$n^{\frac{1}{2}} \left\{ \widehat{\mathrm{ROC}}_t(u) - \mathrm{ROC}_t(u) \right\} = \widehat{\mathscr{W}_{\mathrm{TPR}_t}} \left\{ \mathrm{FPR}_t^{-1}(u) \right\} - \mathrm{ROC}_t(u)\, \widehat{\mathscr{W}_{\mathrm{FPR}_t}} \left\{ \mathrm{FPR}_t^{-1}(u) \right\} + o_p(1),$$

where $\dot{\mathrm{ROC}}_t(u) = \partial \mathrm{ROC}_t(u) / \partial u$. It follows that $n^{\frac{1}{2}} \left\{ \widehat{\mathrm{ROC}}_t(u) - \mathrm{ROC}_t(u) \right\}$ converges weakly to a zero-mean Gaussian process.

REFERENCES

Bang H, Tsiatis A. Estimating Medical Costs with Censored Data. Biometrika. 2000; 87:329–343.

Bang H, Tsiatis A. Median regression with censored cost data. Biometrics. 2002; 58:643–649. [PubMed: 12229999]

Beran, R. Nonparametric regression with randomly censored survival data. Univ. of California; Berkeley: 1981. Unpublished manuscript

Bickel PJ, Rosenblatt M. On some global measures of the deviations of density function estimates (Corr: V3 p1370). Ann. Statist. 1973; 1:1071–1095.

Bilias Y, Gu M, Ying Z. Towards a general asymptotic theory for Cox model with staggered entry. Ann. Statist. 1997; 25:662–682.

Breslow N, Lubin J, Marek P, Langholz B. Multiplicative models and cohort analysis. J. Am. Statist. Assoc. 1983; 78:1–12.

Breslow N, Wellner J. Weighted Likelihood for Semiparametric Models and Two-phase Stratified Samples, with Application to Cox Regression. Scand. J. Statist. 2006; 34:86–102.

Cai G. Almost sure convergence for linear process generated by asymptotically linear negative quadrant dependence processes. Commun. Korean Math. Soc. 2005; 20:161–168.

Cai T, Pepe M, Zheng Y, Lumley T, Jenny N. The sensitivity and specificity of markers for event times. Biostatistics. 2006; 7:182–97. [PubMed: 16079162]

Chen K. Generalized case-cohort sampling. *J. R. Statist. Soc.* B. 2001; 63:791–809.

Colditz G, Manson J, Hankinson S. The Nurses' Health Study: 20-year contribution to the understanding of health among women. Journal of Women's Health. 1997; 6:49–62.

Cook N, Buring J, Ridker P. The effect of including C-reactive protein in cardiovascular risk prediction models for women. Ann. Intern. Med. 2006; 145:21. [PubMed: 16818925]

Cox D. Regression models and life-tables. *J. R. Statist. Soc.* B. 1972:187–220.

Dabrowska D. Uniform consistency of the kernel conditional Kaplan-Meier estimate. Ann. Statist. 1989; 17:1157–67.

Du Y, Akritas M. IID representations of the conditional Kaplan-Meier process for arbitrary distributions. Math. Meth. Statist. 2002; 11:152–82.

Food and Drug Administration. Test determines risk of breast cancer returning. 2007. http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm048477.htm

Gail M, Brinton L, Byar D, Corle D, Green S, Schairer C, Mulvihill J. Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. JNCI Cancer Spectrum. 1989; 81:1879–86.

Goldstein L, Langholz B. Asymptotic theory for nested case-control sampling in the Cox regression model. Ann. Statist. 1992; 20:1903–28.

Heagerty P, Lumley T, Pepe M. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics. 2000; 56:337–44. [PubMed: 10877287]

Heagerty P, Zheng Y. Survival model predictive accuracy and ROC curves. Bio-metrics. 2005; 61:92–105.

Hunter D, Rimm E, Sacks F, Stampfer M, Colditz G, Litin L, Willett W. Comparison of measures of fatty acid intake by subcutaneous fat aspirate, food frequency questionnaire, and diet records in a free-living population of US men. American Journal of Epidemiology. 1992; 135:418–27. [PubMed: 1550093]

Kalbfleisch, JD.; Prentice, RL. The Statistical Analysis of Failure Time Data. John Wiley & Sons; 2002.

Kannel W, Feinleib M, McNamara P, Garrison R, Castelli W. An investigation of coronary heart disease in families: The Framingham O spring Study. American Journal of Epidemiology. 1979; 110:281–90. [PubMed: 474565]

Léon L, Cai T, Wei L. Robust Inferences For Covariate Effects On Survival Time With Censored Linear Regression Models. Statistics in Biosciences. 2009; 1:1–15.

Nan B, Kalbfleisch J, Yu M. Asymptotic theory for the semiparametric accelerated failure time model with missing data. Ann. Statist. 2009; 37:2351–2376.

Pepe M, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am. J. Epidemiol. 2004; 159:882–90. [PubMed: 15105181]

Pollard, D. Empirical processes: theory and applications. Institute of Mathematical Statistics; 1990.

Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coeffcients when some regressors are not always observed. J. Am. Statist. Assoc. 1994; 89:846–866.

Rundle A, Vineis P, Ahsan H. Design options for molecular epidemiology research within cohort studies. Cancer Epidemiology Biomarkers & Prevention. 2005; 14:1899.

Samuelsen S. A psudolikelihood approach to analysis of nested case-control studies. Biometrika. 1997; 84:379–394.

Tsiatis A. A nonidentifiability aspect of the problem of competing risks. Proceedings of the National Academy of Sciences of the United States of America. 1975; 72:20. [PubMed: 1054494]

Ury H. Effciency of case-control studies with multiple controls per case: continuous or dichotomous data. Biometrics. 1975; 31:643–649. [PubMed: 1100136]

van der Vaart A. Weak convergence of smoothed empirical processes. Scand. J. Statist. 1994; 21:501–4.

Ware J. The limitations of risk factors as prognostic tools. N. Eng. J. Med. 2006; 355:2615.

Wilson P, D'Agostino R, Levy D, Belanger A, Silbershatz H, Kannel W. Prediction of coronary heart disease using risk factor categories. Circulation. 1998; 97:1837–47. [PubMed: 9603539]

Zhang L. A functional central limit theorem for asymptotically negatively dependent random fields. Acta Mathematica Hungarica. 2000; 86:237–259.

Zheng Y, Cai T, Pepe M, Levy W. Time-dependent predictive values of prognostic biomarkers with failure time outcome. J. Am. Statist. Assoc. 2008; 103:362–8.

Table 1

Finite sample performance of the proposed CNA based estimators of accuracy for independent censoring with 1 matched controls under the $\mathbb{F}$- and $\mathbb{B}$-sampling. Shown below are the sample mean (Mean), standard error (SSE), average of the estimated standard error (ASE) and empirical coverage level of the 95% confidence intervals (CovP), multiplied by 100, for the proposed point and interval estimators.

(a) Accuracy Estimates at $c_p = F^{-1}(p)$, for k = 1, 2,3,4.

	Truth	**Bias$^{\mathbb{F}}$**	**Bias$^{\mathbb{B}}$**	**SSE$^{\mathbb{F}}$**	**SSE$^{\mathbb{B}}$**	**ASE$^{\mathbb{F}}$**	**ASE$^{\mathbb{B}}$**	**CovP$^{\mathbb{F}}$**	**CovP$^{\mathbb{B}}$**
$FPR_{t0}(c_{.2})$	79.0	0.1	−0.1	2.5	2.5	2.4	2.5	93.1	94.4
$FPR_{t0}(c_{.4})$	58.2	0.1	0.0	3.0	3.1	3.0	3.0	94.4	93.3
$FPR_{t0}(c_{.6})$	37.7	0.2	0.0	2.9	3.0	2.9	2.9	95.1	93.9
$FPR_{t0}(c_{.8})$	17.8	0.2	0.0	2.3	2.2	2.2	2.2	94.2	93.8
$TPR_{t0}(c_{.2})$	97.1	−0.1	−0.1	1.0	1.0	1.1	1.1	94.1	94.2
$TPR_{t0}(c_{.4})$	90.4	−0.4	−0.4	2.0	1.9	2.0	2.0	96.1	95.8
$TPR_{t0}(c_{.6})$	78.5	−0.7	−0.8	2.8	2.8	2.9	2.9	94.9	95.1
$TPR_{t0}(c_{.8})$	57.3	−1.1	−1.3	3.6	3.6	3.7	3.7	95.3	94.0
$NPV_{t0}(c_{.2})$	99.2	0.0	0.0	0.3	0.3	0.3	0.3	94.6	94.2
$NPV_{t0}(c_{.4})$	98.6	0.0	0.0	0.3	0.3	0.3	0.3	95.2	94.7
$NPV_{t0}(c_{.6})$	98.0	0.0	0.0	0.3	0.3	0.3	0.3	96.7	96.4
$NPV_{t0}(c_{.8}$	97.0	0.0	0.0	0.3	0.3	0.3	0.3	96.0	95.7
$PPV_{t0}(c_{.2})$	6.9	−0.1	−0.1	0.5	0.5	0.5	0.5	93.6	92.1
$PPV_{t0}(c_{.4})$	8.5	−0.2	−0.2	0.7	0.7	0.7	0.7	93.5	92.4
$PPV_{t0}(c_{.6})$	11.1	−0.3	−0.3	1.1	1.0	1.0	1.0	92.7	92.1
$PPV_{t0}(c_{.8})$	16.2	−0.5	−0.4	1.9	1.9	1.9	1.9	92.4	93.2

(b) AUC and Accuracy Estimates at TPR of 0.90.

	Truth	Bias$^{\mathbb{F}}$	Bias$^{\mathbb{B}}$	SSE$^{\mathbb{F}}$	SSE$^{\mathbb{B}}$	ASE$^{\mathbb{F}}$	ASE$^{\mathbb{B}}$	CovP$^{\mathbb{F}}$	CovP$^{\mathbb{B}}$
AUC	78.5	−1.0	−0.9	1.9	1.9	1.9	1.9	94.1	92.8
$FPR_{TPR=0.9}$	57.2	0.7	0.6	4.7	4.7	5.2	5.3	96.0	96.7

(b) AUC and Accuracy Estimates at TPR of 0.90.

$NPV_{TPR=0.9}$	98.6	0.0	0.0	0.2	0.2	0.2	0.2	94.7	93.7
$PPV_{TPR=0.9}$	8.6	−0.2	−0.2	0.9	0.9	0.9	0.9	93.1	93.1

Table 2

Finite sample performance of the proposed CNA based estimators of accuracy for independent censoring with 3 matched controls under the $\mathbb{F}$- and $\mathbb{B}$-sampling. Shown below are the sample mean (Mean), standard error (SSE), average of the estimated standard error (ASE) and empirical coverage level of the 95% confidence intervals (CovP), multiplied by 100, for the proposed point and interval estimators.

(a) Accuracy Estimates at $c_p = F^{-1}(p)$, for k = 1, 2,3,4.

	Truth	$\text{Bias}^{\mathbb{F}}$	$\text{Bias}^{\mathbb{B}}$	$\text{SSE}^{\mathbb{F}}$	$\text{SSE}^{\mathbb{B}}$	$\text{ASE}^{\mathbb{F}}$	$\text{ASE}^{\mathbb{B}}$	$\text{CovP}^{\mathbb{F}}$	$\text{CovP}^{\mathbb{B}}$
$\text{FPR}_{t0}(c_{.2})$	79.0	0.0	0.0	1.5	1.5	1.5	1.5	94.0	93.0
$\text{FPR}_{t0}(c_{.4})$	58.2	0.1	0.0	1.7	1.9	1.8	1.8	95.8	93.4
$\text{FPR}_{t0}(c_{.6})$	37.7	0.1	0.0	1.8	1.8	1.7	1.7	94.8	94.3
$\text{FPR}_{t0}(c_{.8})$	17.8	0.1	0.1	1.3	1.3	1.3	1.3	94.3	94.6
$\text{TPR}_{t0}(c_{.2})$	97.1	−0.1	−0.1	1.0	1.0	1.0	1.0	93.7	93.1
$\text{TPR}_{t0}(c_{.4})$	90.4	−0.3	−0.3	1.8	1.8	1.9	1.9	96.8	95.9
$\text{TPR}_{t0}(c_{.6})$	78.5	−0.6	−0.6	2.5	2.5	2.7	2.7	96.4	95.9
$\text{TPR}_{t0}(c_{.8})$	57.3	−0.9	−0.9	3.3	3.3	3.3	3.3	94.3	94.0
$\text{NPV}_{t0}(c_{.2})$	99.2	0.0	0.0	0.3	0.3	0.3	0.3	93.0	93.0
$\text{NPV}_{t0}(c_{.4})$	98.6	0.0	0.0	0.3	0.3	0.3	0.3	95.6	96.1
$\text{NPV}_{t0}(c_{.6})$	98.0	0.0	0.0	0.3	0.3	0.3	0.3	95.6	95.5
$\text{NPV}_{t0}(c_{.8})$	97.0	0.0	0.0	0.3	0.3	0.3	0.3	93.7	95.2
$\text{PPV}_{t0}(c_{.2})$	6.9	−0.1	−0.1	0.5	0.5	0.5	0.5	92.4	92.9
$\text{PPV}_{t0}(c_{.4})$	8.5	−0.2	−0.2	0.6	0.6	0.6	0.6	92.5	93.2
$\text{PPV}_{t0}(c_{.6})$	11.1	−0.2	−0.2	0.8	0.8	0.8	0.8	94.1	92.7
$\text{PPV}_{t0}(c_{.8})$	16.2	−0.5	−0.5	1.4	1.4	1.5	1.5	92.9	93.0

(b) AUC and Accuracy Estimates at TPR of 0.90.

	Truth	$\text{Bias}^{\mathbb{F}}$	$\text{Bias}^{\mathbb{B}}$	$\text{SSE}^{\mathbb{F}}$	$\text{SSE}^{\mathbb{B}}$	$\text{ASE}^{\mathbb{F}}$	$\text{ASE}^{\mathbb{B}}$	$\text{CovP}^{\mathbb{F}}$	$\text{CovP}^{\mathbb{B}}$
AUC	78.5	−0.8	−0.7	1.6	1.6	1.6	1.6	93.4	93.1
$\text{FPR}_{\text{TPR=0.9}}$	57.2	0.7	0.6	4.2	4.1	4.6	4.7	95.9	96.5

(b) AUC and Accuracy Estimates at TPR of 0.90.

	Truth	Bias^F	Bias^B	SSE^F	SSE^B	ASE^F	ASE^B	CovP^F	CovP^B
$\text{NPV}_{\text{TPR}=0.9}$	98.6	0.0	0.0	0.2	0.2	0.2	0.2	95.3	95.5
$\text{PPV}_{\text{TPR}=0.9}$	8.6	−0.2	−0.2	0.8	0.8	0.8	0.8	93.5	94.5

Table 3

Finite sample performance of the proposed CNA based estimators of accuracy for dependent censoring with 1 matched control under the $\mathbb{F}$- and $\mathbb{B}$-sampling. Shown below are the sample mean (Mean), standard error (SSE), average of the estimated standard error (ASE) and empirical coverage level of the 95% confidence intervals (CovP), multiplied by 100, for the proposed point and interval estimators.

(a) Accuracy Estimates at $c_p = F^{-1}(p)$, for k = 1, 2,3,4.

	Truth	$\text{Bias}^{\mathbb{F}}$	$\text{Bias}^{\mathbb{B}}$	$\text{SSE}^{\mathbb{F}}$	$\text{SSE}^{\mathbb{B}}$	$\text{ASE}^{\mathbb{F}}$	$\text{ASE}^{\mathbb{B}}$	$\text{CovP}^{\mathbb{F}}$	$\text{CovP}^{\mathbb{B}}$
$\text{FPR}_{t0}(c_{.2})$	79.0	0.0	0.0	1.4	1.3	1.4	1.4	94.8	94.7
$\text{FPR}_{t0}(c_{.4})$	58.2	0.1	0.0	1.6	1.5	1.6	1.6	94.7	95.7
$\text{FPR}_{t0}(c_{.6})$	37.7	0.1	0.0	1.5	1.5	1.4	1.4	92.8	94.7
$\text{FPR}_{t0}(c_{.8})$	17.8	0.1	0.1	1.1	1.1	1.1	1.1	94.0	95.2
$\text{TPR}_{t0}(c_{.2})$	97.1	−0.1	−0.1	1.0	1.0	1.0	1.0	94.3	94.5
$\text{TPR}_{t0}(c_{.4})$	90.4	−0.3	−0.3	1.7	1.8	1.8	1.8	95.8	95.5
$\text{TPR}_{t0}(c_{.6})$	78.5	−0.5	−0.5	2.5	2.5	2.5	2.5	95.1	95.1
$\text{TPR}_{t0}(c_{.8})$	57.3	−0.6	−0.6	2.9	2.9	3.0	3.0	95.6	95.5
$\text{NPV}_{t0}(c_{.2})$	99.2	0.0	0.0	0.3	0.3	0.3	0.3	93.0	92.7
$\text{NPV}_{t0}(c_{.4})$	98.6	0.0	0.0	0.3	0.3	0.3	0.3	94.9	94.9
$\text{NPV}_{t0}(c_{.6})$	98.0	0.0	0.0	0.3	0.3	0.3	0.3	95.1	94.5
$\text{NPV}_{t0}(c_{.8})$	97.0	0.0	0.0	0.3	0.3	0.3	0.3	95.2	94.9
$\text{PPV}_{t0}(c_{.2})$	6.9	−0.1	−0.1	0.4	0.4	0.4	0.4	94.1	93.1
$\text{PPV}_{t0}(c_{.4})$	8.5	−0.2	−0.2	0.5	0.5	0.5	0.5	93.1	93.3
$\text{PPV}_{t0}(c_{.6})$	11.1	−0.3	−0.2	0.8	0.8	0.8	0.8	92.4	93.4
$\text{PPV}_{t0}(c_{.8})$	16.2	−0.4	−0.4	1.3	1.3	1.3	1.3	94.1	94.6

(b) AUC and Accuracy Estimates at TPR of 0.90.

	Truth	$\text{Bias}^{\mathbb{F}}$	$\text{Bias}^{\mathbb{B}}$	$\text{SSE}^{\mathbb{F}}$	$\text{SSE}^{\mathbb{B}}$	$\text{ASE}^{\mathbb{F}}$	$\text{ASE}^{\mathbb{B}}$	$\text{CovP}^{\mathbb{F}}$	$\text{CovP}^{\mathbb{B}}$
AUC	78.5	−0.7	−0.7	1.5	1.5	1.5	1.5	93.0	92.0
$\text{FPR}_{\text{TPR=0.9}}$	57.2	0.6	0.6	4.1	4.2	4.5	4.5	96.2	95.9

(b) AUC and Accuracy Estimates at TPR of 0.90.

	Truth	Bias^F	Bias^B	SSE^F	SSE^B	ASE^F	ASE^B	CovP^F	CovP^B
$\text{NPV}_{\text{TPR}=0.9}$	98.6	0.0	0.0	0.2	0.2	0.2	0.2	94.8	93.5
$\text{PPV}_{\text{TPR}=0.9}$	8.6	−0.2	−0.2	0.8	0.8	0.8	0.8	93.9	93.8

J Am Stat Assoc. Author manuscript; available in PMC 2012 July 25.

Table 4

Comparing the performance of the proposed CNA and DIPW methods under the $\mathbb{F}$- and $\mathbb{B}$-sampling. Shown below are the percent of relative bias relative and the root mean squared error (MSE) (multiplied by 100) for accuracy estimators with 3 matched controls.

(a) Independent Censoring

		Percent of Relative Bias				100× Root MSE			
	Truth	**DIPW**$^{\mathbb{F}}$	**CNA**$^{\mathbb{F}}$	**DIPW**$^{\mathbb{B}}$	**CNA**$^{\mathbb{B}}$	**DIPW**$^{\mathbb{F}}$	**CNA**$^{\mathbb{F}}$	**DIPW**$^{\mathbb{B}}$	**CNA**$^{\mathbb{B}}$
$\text{FPR}_{t0}(c_{.2})$	79.0	0.0	0.0	0.0	0.1	1.3	1.5	1.3	1.5
$\text{FPR}_{t0}(c_{.8})$	17.8	0.1	0.7	0.4	0.5	1.1	1.4	1.7	1.3
$\text{TPR}_{t0}(c_{.2})$	97.1	0.0	−0.1	0.0	−0.1	1.0	1.0	1.1	1.0
$\text{TPR}_{t0}(c_{.8}$	57.3	−0.2	−1.6	−0.2	−1.6	3.1	3.4	4.0	3.4
$\text{NPV}_{t0}(c_{.2})$	99.2	0.0	0.0	0.0	0.0	0.3	0.3	0.3	0.3
$\text{NPV}_{t0}(c_{.8})$	97.0	0.0	0.0	0.0	0.0	0.3	0.3	0.3	0.3
$\text{PPV}_{t0}(c_{.2})$	6.9	−0.1	−1.6	−0.1	−1.7	0.4	0.5	0.4	0.5
$\text{PPV}_{t0}(c_{.8})$	16.2	−0.1	−3.0	−0.2	−2.9	1.4	1.5	1.5	1.5
AUC	78.5	−0.4	−1.0	−0.4	−0.9	1.6	1.8	1.6	1.8
$\text{FPR}_{\text{TPR}=0.9}$	57.2	−1.7	1.2	−1.7	1.1	4.6	4.3	4.6	4.1
$\text{NPV}_{\text{TPR}=0.9}$	98.6	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.2
$\text{PPV}_{\text{TPR}=0.9}$	8.6	1.7	−2.1	1.7	−2.2	0.9	0.8	0.9	0.8

(b) Dependent Censoring

		Percent of Relative Bias				100× Root MSE			
	Truth	**DIPW**$^{\mathbb{F}}$	**CNA**$^{\mathbb{F}}$	**DIPW**$^{\mathbb{B}}$	**CNA**$^{\mathbb{B}}$	**DIPW**$^{\mathbb{F}}$	**CNA**$^{\mathbb{F}}$	**DIPW**$^{\mathbb{B}}$	**CNA**$^{\mathbb{B}}$
$\text{FPR}_{t0}(c_{.2})$	79.0	5.4	−0.0	5.5	0.1	4.4	1.4	4.5	1.3
$\text{FPR}_{t0}(c_{.8})$	17.8	15.2	0.4	15.6	0.3	3.0	1.1	3.4	1.1
$\text{TPR}_{t0}(c_{.2})$	97.1	0.2	−0.1	0.2	−0.1	1.0	1.0	1.0	1.0
$\text{TPR}_{t0}(c_{.8}$	57.3	1.3	−1.0	1.2	−1.0	3.2	3.0	4.0	3.0
$\text{NPV}_{t0}(c_{.2})$	99.2	−0.2	−0.0	−0.2	−0.0	0.4	0.3	0.4	0.3

(b) Dependent Censoring

		Percent of Relative Bias				100× Root MSE			
	Truth	**DIPW**^{F}	**CNA**^{F}	**DIPW**^{B}	**CNA**^{B}	**DIPW**^{F}	**CNA**^{F}	**DIPW**^{B}	**CNA**^{B}
$NPV_{t0}(c_{.8})$	97.0	−0.1	0.0	−0.1	0.0	0.3	0.3	0.4	0.3
$PPV_{t0}(c_{.2})$	6.9	−2.5	−1.6	−2.5	−1.7	0.4	0.4	0.5	0.4
$PPV_{t0}(c_{.8})$	16.2	−8.4	−2.4	−8.6	−2.3	1.8	1.3	1.9	1.3
AUC	78.5	−2.8	−0.9	−2.8	−0.9	3.0	2.0	3.0	2.0
$FPR_{TPR=.9}$	57.2	6.0	1.0	6.0	1.0	6.0	4.0	6.0	4.0
$NPV_{TPR=.9}$	98.6	−0.2	−0.0	−0.2	−0.0	0.0	0.0	0.0	0.0
$PPV_{TPR=.9}$	8.6	−2.9	−2.0	−2.9	−2.0	1.0	0.9	1.0	0.9

Table 5

Estimated accuracy summaries of a risk score for predicting 5-year survival from the Framingham Offspring Study. Shown below are estimates (Est) and standard errors (SE) of accuracy estimators (multiplied by 100) based on the CNA and DIPW methods under the $\mathbb{F}$-sampling with full cohort data and nested case-control samples using $m = 1$ or 3 matched controls. Here c_p is the pth percentile of the observed risk score in the full cohort.

	CNA						DIPW					
	Cohort		1 control		3 controls		Cohort		1 control		3 controls	
	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE	Est	SE
$FPR_5(c_{.2})$	79.7	1.0	79.7	2.5	79.6	1.6	79.7	1.0	79.6	2.5	79.8	1.5
$FPR_5(c_{.4})$	59.4	1.2	59.3	3.1	59.4	2.0	59.2	1.2	59.0	3.0	59.3	1.9
$FPR_5(c_{.6})$	39.2	1.2	39.2	3.1	39.0	2.0	38.8	1.3	38.6	3.0	38.9	1.9
$FPR_5(c_{.8})$	19.1	1.0	19.1	2.6	19.1	1.7	18.8	0.9	18.8	2.4	18.8	1.5
$TPR_5(c_{.2})$	92.8	4.5	93.4	4.6	93.2	4.6	91.9	4.3	91.9	4.5	91.9	4.5
$TPR_5(c_{.4})$	86.9	5.1	87.0	5.2	87.0	5.1	89.2	4.7	89.2	5.1	89.2	5.1
$TPR_5(c_{.6})$	78.5	6.8	78.3	6.9	78.4	6.8	78.4	6.5	78.4	6.8	78.4	6.8
$TPR_5(c_{.8})$	61.2	7.9	60.0	8.6	60.6	8.0	62.2	7.7	62.2	8.0	62.2	8.0
$NPV_5(c_{.2})$	99.2	0.5	99.3	0.5	99.3	0.5	99.1	0.5	99.1	0.5	99.1	0.5
$NPV_5(c_{.4})$	99.3	0.3	99.3	0.3	99.3	0.3	99.4	0.3	99.4	0.3	99.4	0.3
$NPV_5(c_{.6})$	99.2	0.3	99.2	0.3	99.2	0.3	99.2	0.3	99.2	0.3	99.2	0.3
$NPV_5(c_{.8})$	99.0	0.3	98.9	0.3	99.0	0.3	99.0	0.3	99.0	0.3	99.0	0.3
$PPV_5(c_{.2})$	2.5	0.4	2.5	0.4	2.5	0.4	2.5	0.4	2.5	0.4	2.5	0.4
$PPV_5(c_{.4})$	3.1	0.5	3.1	0.6	3.1	0.5	3.2	0.6	3.2	0.6	3.2	0.6
$PPV_5(c_{.6})$	4.2	0.7	4.2	0.9	4.2	0.8	4.2	0.8	4.3	0.9	4.2	0.8
$PPV_5(c_{.8})$	6.5	1.3	6.4	1.7	6.5	1.4	6.8	1.4	6.9	1.6	6.8	1.4
AUC	75.2	4.1	74.9	4.5	75.1	4.2	75.8	3.9	75.5	4.3	75.7	4.2
$FPR_{TPR=.9}$	65.0	13.9	66.3	14.5	65.8	13.8	58.7	8.4	58.5	14.2	58.8	12.5
$PPV_{TPR=.9}$	99.4	0.3	99.4	0.3	99.4	0.3	99.4	0.7	99.4	1.0	99.4	0.9
$NPV_{TPR=.9}$	2.9	0.8	2.8	0.8	2.9	0.8	3.2	0.2	3.3	0.2	3.2	0.2

J Am Stat Assoc. Author manuscript; available in PMC 2012 July 25.

Table 6

Finite sample performance of the proposed CNA based estimators of accuracy for independent censoring with 1 matched controls under both the $\mathbb{F}$- and $\mathbb{B}$-sampling for $n = 2000$. Shown below are the sample mean (Mean), standard error (SSE), average of the estimated standard error (ASE) and empirical coverage level of the 95% confidence intervals (CovP), multiplied by 100, for the proposed point and interval estimators of the accuracy measures at $c_p = \mathscr{F}^{-1}(p)$

(a).

	Truth	Bias^F	Bias^B	SSE^F	SSE^B	ASE^B	ASE^B	CovP^F	CovP^B
$FPR_{t0}(c_{.2})$	77.4	0.1	0.1	2.6	2.7	2.6	2.6	94.0	94.7
$FPR_{t0}(c_{.4})$	55.4	0.1	0.2	3.1	3.1	3.1	3.1	94.4	94.8
$FPR_{t0}(c_{.6})$	34.3	0.1	0.1	2.8	2.8	2.9	2.9	95.8	96.5
$FPR_{t0}(c_{.8})$	14.7	0.3	0.3	2.1	2.1	2.1	2.1	95.2	95.0
$TPR_{t0}(c_{.2})$	96.8	−0.2	−0.2	1.1	1.1	1.2	1.2	95.5	95.2
$TPR_{t0}(c_{.4})$	89.3	−0.5	−0.5	1.9	2.0	2.1	2.1	96.9	96.5
$TPR_{t0}(c_{.6})$	76.2	−0.8	−0.8	2.8	2.8	3.0	3.0	96.2	95.9
$TPR_{t0}(c_{.8})$	53.6	−0.9	−0.1	3.6	3.5	3.6	3.7	94.3	94.9
$NPV_{t0}(c_{.2})$	97.8	−0.1	−0.1	0.8	0.7	0.8	0.8	95.2	94.4
$NPV_{t0}(c_{.4})$	96.4	−0.1	−0.1	0.7	0.7	0.7	0.7	96.7	95.7
$NPV_{t0}(c_{.6})$	94.6	−0.1	−0.1	0.7	0.7	0.7	0.7	95.8	95.4
$NPV_{t0}(c_{.8})$	92.1	−0.1	−0.1	0.8	0.8	0.8	0.8	95.6	95.5
$PPV_{t0}(c_{.2})$	16.4	−0.2	−0.2	1.2	1.1	1.2	1.2	93.7	94.2
$PPV_{t0}(c_{.4})$	20.2	−0.3	−0.3	1.5	1.5	1.5	1.5	95.0	94.0
$PPV_{t0}(c_{.6})$	25.9	−0.5	−0.5	2.1	2.1	2.2	2.2	95.8	93.9
$PPV_{t0}(c_{.8})$	36.4	−0.1	−0.1	3.4	3.5	3.6	3.6	95.2	93.8