Published in final edited form as:

J Am Stat Assoc. 2018 ; 113(522): 882–892. doi:10.1080/01621459.2017.1295866.

Nonparametric Maximum Likelihood Estimators of Time-Dependent Accuracy Measures for Survival Outcome Under Two-Stage Sampling Designs

Dandan Liu^a [Assistant Professor], Tianxi Cai^b [Full Professor], Anna Lok^c [Full Professor], and Yingye Zheng^{d,†} [Full Members]

^aDepartment of Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37232

^bDepartment of Biostatistics, Harvard School of Public Health, Boston, Massachusetts 02115

^cDivision of Gastroenterology, University of Michigan, Ann Arbor, MI 48109

^dPublic Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109

Abstract

Large prospective cohort studies of rare chronic diseases require thoughtful planning of study designs, especially for biomarker studies when measurements are based on stored tissue or blood specimens. Two-phase designs, including nested case-control (Thomas, 1977) and case-cohort (Prentice, 1986) sampling designs, provide cost-effective strategies for conducting biomarker evaluation studies. Existing literature for biomarker assessment under two-phase designs largely focuses on simple inverse probability weighting (IPW) estimators (Cai and Zheng, 2011; Liu et al., 2012). Drawing on recent theoretical development on the maximum likelihood estimators for relative risk parameters in two-phase studies (Scheike and Martinussen, 2004; Zeng et al., 2006), we propose nonparametric maximum likelihood based estimators to evaluate the accuracy and predictiveness of a risk prediction biomarker under both types of two-phase designs. In addition, hybrid estimators that combine IPW estimators and maximum likelihood estimation procedure are proposed to improve efficiency and alleviate computational burden. We derive large sample properties of proposed estimators and evaluate their finite sample performance using numerical studies. We illustrate new procedures using a two-phase biomarker study aiming to evaluate the accuracy of a novel biomarker, des- γ -carboxy prothrombin, for early detection of hepatocellular carcinoma (Lok et al., 2010).

Keywords

Case-cohort sampling; Negative predictive value; Nested case-control sampling; Nonparametric maximum likelihood estimator; Positive predictive value; Receiver Operating Characteristics Curve (ROC curve); Two-phase study

[†]Corresponding author (yzheng@fhcrc.org).

SUPPLEMENTARY MATERIAL

Supplementary material includes regularity conditions, proof of asymptotic results of the proposed estimators, and further simulation results.

1. INTRODUCTION

Large prospective cohort studies of rare chronic diseases such as cancer often require careful consideration of many aspects of study designs, from selection of study subjects, measurement of key variables and auxiliary information, to the analysis of results. Thoughtful planning is especially demanding in cancer biomarker research, when measurements are based on stored tissue or blood specimens. Considering the labor and costs associated with measuring biomarkers, two major design choices, the nested case-control design (Thomas, 1977; Prentice and Breslow, 1978), and the case-cohort design (Prentice, 1986), have often been considered for valid and cost-efficient sampling of individuals from a cohort for more expensive measurements. Both designs are conceptually conducted in two phases: in the first phase routine clinical risk factors are collected on each individual in the cohort; in the second phase, more expensive exposures such as biomarker measurements are then ascertained on individuals who developed rare event of interest and a subset of individuals without events. The two designs, originally developed in epidemiological association studies, have been adopted for risk marker evaluation (Lok et al., 2010; Folsom et al., 2002; Wang et al., 2011).

There is vast literature on statistical analyses incorporating two phase designs, focusing mainly on making inference on relative or absolute risk parameters. In matched nested case-control studies, the classic approach to estimating hazard ratio parameters is to fit a conditional logistic regression model, or equivalently the proportional hazards model (Goldstein and Langholz, 1992). The asymptotic properties have been formally derived for the regression parameters (Goldstein and Langholz, 1992) and absolute risk (Langholz and Borgan, 1997) under the partial likelihood in a proportional hazards model using counting process and martingale theory (Andersen and Gill, 1982). For case-cohort studies, estimating regression parameters in a Cox model is based on a pseudo-likelihood approach under the original design (Prentice, 1986), and extending to more flexible exposure stratified case-cohort sampling is considered in Borgan et al. (2000). In recent years, analyses of two-phase designs have been cast in the missing data framework, appealing to the theory of empirical process (Pollard, 1990). One popular approach is based on the idea of inverse probability weighted (IPW) estimators, where individual contributions are weighted inversely proportional to their sampling fractions (Lin, 2000; Breslow and Wellner, 2007; Gray, 2009). In particular, Breslow and Wellner (2007) derived key theoretical results for estimates of hazard ratios and cumulative event probabilities in the population under Cox regression with stratified case-cohort sampling. Adapting IPW estimators to nested case-control studies has also been proposed (Samuelsen, 1997; Chen, 2001; Cai and Zheng, 2012). The other approach is the consideration of nonparametric maximum likelihood estimators (NPMLE), which attain the semiparametric efficiency bound and are therefore potentially more efficient than the IPW estimators. For example, Scheike and Martinussen (2004) and Scheike and Juul (2004) proposed a maximum likelihood estimator for both designs, and Zeng et al. (2006) and Zeng and Lin (2014) extended the maximum likelihood estimators to the more general class of semiparametric transformation models. To date, such NPMLE based estimation method has not been developed for evaluation of accuracy measures of risk prediction biomarkers.

In biomarker research, the inference goal is often to characterize the clinical utility of a novel biomarker in predicting important clinical outcomes of interest. A consensus in the field is that relative risks or absolute risks are not sufficient for the purpose of evaluating the clinical utility of risk markers. Specifically, for a putative marker or a predictive score Y measured at baseline and used to predict T , the time to an event of interest, by a future time t , two classes of time-dependent accuracy parameters, retrospective and prospective accuracies, are commonly calculated as summaries of discrimination. The retrospective accuracies, including the true and false positive rate (TPR and FPR) functions, evaluated at a future predicting time t , and a threshold c , are defined as

$$\text{TPR}_t(c) = \text{pr}(Y > c | T \leq t), \quad \text{and} \quad \text{FPR}_t(c) = \text{pr}(Y > c | T > t);$$

The time-dependent receiver operating characteristic (ROC) curve is a plot of $\text{TPR}_t(c)$ versus $\text{FPR}_t(c)$, across all possible values of c , and the time-dependent area under the ROC curve (AUC) summarizes the discriminatory capacity of the marker. The mean risks in cases and controls are defined as

$$\text{ITP}_t = \int_c \text{TPR}_t(c) dF(c), \quad \text{and} \quad \text{IFP}_t = \int_c \text{FPR}_t(c) dF(c);$$

where $R(c) = \text{pr}(Y \leq c)$. The difference in mean risks at t , $\text{DMR}_t = \text{ITP}_t - \text{IFP}_t$, also provides a summary for the performance of the marker. The prospective accuracy summaries, positive and negative predictive values (PPV and NPV), are time-dependent functions of the form

$$\text{PPV}_t(c) = \text{pr}(T \leq t | Y > c), \quad \text{and} \quad \text{NPV}_t(c) = \text{pr}(T > t | Y \leq c).$$

Clearly, estimation of these summary measures of predictive performance of risk markers adds another level of complexity, requiring additional estimation of the distribution of risk in the population and its related functionals. Extending the existing statistical literature to make efficient inference about these predictive accuracy parameters under two-phase studies is urgently needed. Recently novel statistical tools for making inference about the predictive performance of a novel biomarker have also been developed for both matched nested case-control and stratified case-cohort studies (Cai and Zheng, 2012; Liu et al., 2012; Cai and Zheng, 2011). In these approaches, simple IPW estimators were considered, with selection probabilities reflecting the sampling scheme of a specific design. While such an IPW estimation procedure is simple to implement, several limitations exist. When true selection probabilities are used in the simple IPW estimators, the method may yield quite inefficient estimators, because it discards information from individuals with incomplete data. Especially in the context of studying rare diseases with only a limited number of events, improving analytical efficiency is of great practical importance. Furthermore, in many, perhaps in most, practical situations the two-phase sampling plans can be quite complicated, due to practical considerations such as the need to reuse samples that were previously assayed in other studies, or mismeasurement due to inadequate samples. Retrieving ‘true’ sampling weights in such study settings can be considerably difficult. Developing more efficient yet

more flexible estimating procedures for validating risk models under complex cohort sampling is therefore desirable.

In this manuscript we propose three efficient estimators of predictive accuracy summaries to improve efficiency over the existing IPW estimators (Cai and Zheng, 2011; Liu et al., 2012) in the context of biomarker assessment. Some important developments are considered in this manuscript. First, we extend the literature of relative risk NPMLE estimators (Scheike and Martinussen, 2004; Zeng et al., 2006) to the estimation of an array of summary indices of the clinical performances of biomarkers. In particular we provide further estimating steps in order to calculate our summary measures of interests. We formally develop asymptotic theory for the NPMLE estimators for predictive performance summaries, drawing on the literature on empirical theory for NPMLE relative risk estimators (Zeng and Lin, 2014). Second, since the estimators require additional steps for estimating explicitly the empirical conditional distribution of biomarkers given other matching/stratification covariates, the well known issue of convergence with a NPMLE estimator becomes more salient. In situations with convergence issue, we propose to consider a class of hybrid likelihood based estimators that combine IPW estimators with the maximum likelihood estimation procedure, to both improve efficiency over existing IPW estimators and reduce computational steps required in the fitting. The corresponding hybrid estimation of summary measures directly follows. In addition, we consider best linear unbiased estimators using the proposed hybrid estimators and IPW estimators to guarantee improvement on efficiency over IPW estimators. The theoretical property of all three types of estimators are also investigated.

The paper is organized as follows: In Section 2 we define notations and specify general models for outcome T and biomarker Y . In Section 3 we propose estimation procedures for both regular NPMLE estimators, one-step estimators and best linear unbiased estimators (BLUE) combining IPW estimators and the one-step estimators. Large sample properties of the proposed estimators are described in Section 4. Section 5 illustrated how the methods can be extended to the evaluation of multiple markers for predicting the event times. Numerical evaluations of the proposed point and interval estimators and comparison to existing estimators via simulation studies are presented in Section 6. We illustrate the proposed methods with a nested case-control study aimed to assess the predictive performance of a novel biomarker for liver cancer early detection in Section 7. Final remarks are given in Section 8.

2. NOTATION AND SETTINGS

Let T , C and $X = \min(T, C)$ denote the failure, censoring and follow-up times respectively. The event indicator is $\delta = I(T < C)$. Let Y , with finite support $\mathcal{B}_Y = [c_l, c_u]$, denote the biomarker measurement whose clinical utility is to be evaluated. Under a two-phase sampling design, Y is measured only for a subset of subjects selected into a phase II subcohort. We consider here a general framework in which selection at the second phase is also dependent on a stratification variable Z , which is assumed to take K unique values, $z_1, \dots, z_K$. If Z is a constant, this design reduces to the conventional two-phase sampling design. In a stratified case-cohort design, the subset is randomly selected from each of the strata defined by δ and Z . Under a matched nested case-control sampling design, all individuals

observed to have an event with $\delta = 1$ are selected as *cases* for further evaluation of marker Y . For each of the observed cases with failure time t and $Z = z$, m controls are selected without replacement from the matched risk set $\{j: X_j \leq t, Z_j = z\}$. Suppose the full data consists of n independent and identically distributed random vectors,

$\mathbb{F} = \{\mathcal{F}_i = (X_i, \delta_i, Z_i, Y_i), i = 1, \dots, n\}$ and the observable data consists of

$\mathbb{O} = \{\mathcal{O}_i = (X_i, \delta_i, Z_i, V_i, Y_i), i = 1, \dots, n\}$.

To calculate our parameter of interests including $\text{TPR}_t(c)$, $\text{FPR}_t(c)$, $\text{PPV}_t(c)$ and $\text{NPV}_t(c)$, we notice that with simple derivation using Bayes' theorem, the aforementioned accuracy measures for Y can be expressed as functions of the above three quantities as

$$\text{TPR}_t(c) = \frac{\{1 - F(c)\} - S(t, c)}{1 - S_T(t)}, \quad \text{FPR}_t(c) = \frac{S(t, c)}{S_T(t)}, \quad (1)$$

$$\text{PPV}_t(c) = \frac{\{1 - F(c)\} - S(t, c)}{1 - F(c)}, \quad \text{NPV}_t(c) = \frac{S_T(t) - S(t, c)}{F(c)},$$

where $S_T(t) = P(T > t)$. Many other accuracy measures are functions of the above measures.

For example, $\text{ROC}_t(u) = \text{TPR}_t\{\text{FPR}_t^{-1}(u)\}$, $\text{AUC}_t = \int_0^1 \text{ROC}_t(u) du$ and $\text{DMR}_t = \int$

$\text{TPR}_t(c) dF(c) - \int \text{FPR}_t(c) dF(c)$. Therefore the construction of all the accuracy measures involves only $S(t, c)$, since $S_T(t) = \Pr(T > t, Y > c) = S(t, c)$ and $F(c) = 1 - \Pr(Y > c) = 1 - S(0, c)$.

To facilitate the estimation of $S(t, c)$, we specify that the relationship between T_i and Y_i follows a semiparametric transformation model $\Pr(T_i \leq t | Y_i) = 1 - \mathcal{T}_0[\log\{\Lambda(t)\} + \beta Y]$ (Cheng et al., 1995, 1997; Zeng and Lin, 2006), where $\mathcal{T}_0$ is some distributional function, $\Lambda(t)$ is an unspecified increasing function, and β is a regression parameter. For simplicity, we present results only for $\mathcal{T}_0(x) = \exp\{-\exp(x)\}$, which leads to the proportional hazards

model (Cox, 1972), with $\Lambda(t) = \int_0^t \lambda(s) ds$, where $\lambda(t)$ is the baseline hazard at t . The

corresponding conditional hazard and survival functions are $\lambda_T(t | Y = y) = \lambda(t) \exp(\beta y)$ and $S_T(t | Y = y) = \exp\{-\Lambda(t) \exp(\beta y)\}$. Let $G(\cdot | \cdot)$ denote the conditional cumulative distribution function of Y given Z . Note that $F(\cdot) = \sum_{k=1}^K G(\cdot | z_k) p_k$, where p_k is the

probability mass at z_k . We further let $S(t, c) = \int_{c_l}^c S_T(t | y) dF(y)$ denote the bivariate survival

function of Y and T .

We have now clarified the clinical setting such a model might be useful in Section 2. We also commented in Section 2 on model specification that 'Note that we assume here $\Pr(T_i | Y_i, Z_i) = \Pr(T_i | Y_i)$, that is, the time to clinical event depends on the stratum only through the predictor Y . Such an assumption is often reasonable in the *validation* study setting we

considered. For example, when Y represents a composite linear score of the form $Y = \sum_{k=1}^K w_k \mathcal{Y}_k$ such as the Reynolds risk score (Cook et al., 2012) and Z is one of the routine risk factors $\{\mathcal{Y}_k, k = 1, \dots, K\}$, it is reasonable to assume that T only depends on $\{\mathcal{Y}_1, \dots, \mathcal{Y}_K\}$ through Y and thus is independent of Z given Y . As another example, if Y is a novel marker measured with an expensive technology while Z is a less accurate version of Y ascertained using a routine procedure, then it is also reasonable to assume that T is independent of Y given Z . We describe a setting where predictors including both a new marker Y and other existing prognostic factors in Section 5.

3. ESTIMATION

3.1 NPMLE of Accuracy Summaries

Two standard assumptions are required for the proposed procedures: (i) independent censoring assumption; and (ii) missing-at-random assumption. The independent censoring assumption assumes that the censoring time C is independent of the failure time T given Y and Z among selected subcohort and given Z among unselected cohort; Further, we make an additional assumption that the censoring is noninformative on Y given Z . The missing-at-random assumes that V does not depend on Y given the observed information in the first phase, i.e. (X, δ, Z) , and thus implies that the sampling probabilities could be omitted from the likelihood function. Similar arguments were made in Scheike and Juul (2004) and Zeng and Lin (2014). Therefore, the observed likelihood function of the parameter of interest $\theta = \{\beta, \Lambda, G\}$ is then

$$\begin{aligned} \mathcal{L}_n(\theta) = & \prod_{i=1}^n \left\{ \lambda_T(X_i | Y_i)^{\delta_i} S_T(X_i | Y_i) dG(Y_i | Z_i) \right\}^{V_i} \quad (2) \\ & \times \left\{ \int_{c_l}^{c_u} \lambda_T(X_i | c)^{\delta_i} S_T(X_i | c) dG(c | Z_i) \right\}^{1-V_i}. \end{aligned}$$

NPMLEs $\hat{\theta} = \{\hat{\beta}, \hat{\Lambda}, \hat{G}\}$ for θ can be obtained by iteratively maximizing the likelihood function (2) as detailed in Chen (2002), Scheike and Martinussen (2004) and Scheike and Juul (2004). The NPMLE estimators for $F(c)$ and $S(t, c)$ could be obtained by plugging in $\hat{\theta}$ as $\hat{F}(c) = \sum_{k=1}^K \hat{G}(c | z_k) \hat{p}_k$ and $\hat{S}(t, c) = \int_c^{c_u} \hat{S}_T(t | y) d\hat{F}(y)$ where $\hat{p}_k$ is the empirical estimator for p_k and $\hat{S}_T(t | y) = \exp\{-\hat{\Lambda}(t) \exp(\hat{\beta}_y)\}$. Subsequently, the plug-in NPMLE estimators of $\text{TPR}_\ell(c)$, $\text{FPR}_\ell(c)$, $\text{PPV}_\ell(c)$ and $\text{NPV}_\ell(c)$, $\widehat{\text{TPR}}_\ell(c)$, $\widehat{\text{FPR}}_\ell(c)$, $\widehat{\text{PPV}}_\ell(c)$ and $\widehat{\text{NPV}}_\ell(c)$ respectively, can be easily calculated using Equations (1) and NPMLE estimator $\hat{S}(t, c)$.

3.2 Hybrid one-step Estimators

To numerically solve for the NPMLE, expectation-maximization algorithms are often adopted (Scheike and Martinussen, 2004). It is well recognized that these algorithms often have convergence issues – either it requires long iterations or it fails to converge. These issues are especially pronounced when the dimension of the parameter space is large as is the proposed procedure where both the baseline hazard function and the distribution of the biomarkers are specified nonparametrically. To overcome such computational difficulty, we propose one-step expectation-maximization hybrid estimators that take advantage of a consistent initial estimator. Similar concepts were considered in Robinson (1988) to study the stochastic difference of two consistent estimators where one is obtained using iterative methods (e.g. expectation-maximization and Newton-Raphson method) with the other one as the initial values. They also showed that a one-step Newton-Raphson estimator using a $n^{1/2}$ -consistent initial estimator is also $n^{1/2}$ -consistent. However, such an approach does not work well here since the Newton-Raphson is computationally challenging for the NPMLE. On the other hand, the proposed one-step expectation-maximization estimator could be implemented using standard statistical software and the standard error could be obtained using bootstrap without much computational burden. We next detail the one-step expectation-maximization estimator.

To start we first obtain the IPW estimator of $\beta, \tilde{\beta}$, by maximizing a weighted partial log-likelihood with score equation

$$\tilde{U}_{\text{IPW}}(\beta) = \sum_{i=1}^n \omega_i \delta_i \left\{ Y - \frac{\sum_{j=1}^n \omega_j I(X_j \geq X_i) Y_j \exp(\beta Y_j)}{\sum_{j=1}^n \omega_j I(X_j \geq X_i) \exp(\beta Y_j)} \right\}.$$

A simple estimator of ω_i , $\tilde{\omega}_i = V_i / \tilde{\pi}_i$, can be derived based on the sampling scheme of the study. For a stratified case-cohort design, Liu et al. (2012) considered sampling probability of

$$\tilde{\pi}_i(\delta, Z) = \sum_{d=0}^1 \frac{I(\delta_i = d) \sum_j I(\delta_j = d) I(V_j = 1, Z_j = Z_i)}{\sum_j I(\delta_j = d) I(Z_j = Z_i)}.$$

For a matched nested case-control design with matching covariates Z and matching ratio m , $\tilde{\pi}_i$ can be calculated as

$$\tilde{\pi}_i(\delta, X, Z) = \delta_i + (1 - \delta_i) \left[1 - \prod_{j: X_j < X_i} \left\{ 1 - \frac{m \delta_j I(Z_j = Z_i)}{\sum_k (X_k \geq X_j) - 1} \right\} \right]$$

(Cai and Zheng, 2011). An IPW estimator of $\Lambda_0(t)$ can be derived as

$$\tilde{\Lambda}_0(t) = \sum_{i=1}^N \frac{\tilde{\omega}_i I(X_i \leq t) \delta_i}{\sum_{j=1}^n \tilde{\omega}_j I(X_j \geq X_i) \exp(\tilde{\beta} Y_j)},$$

and an IPW empirical estimator for $G(c/z)$ is

$$\tilde{G}(c|z) = \sum_{i=1}^N \tilde{\omega}_i I(Y_i < c, Z_i = z) / \left\{ \sum_i \tilde{\omega}_i I(Z_i = z) \right\}$$

With IPW estimator $\tilde{\theta} = (\tilde{\beta}, \tilde{\Lambda}, \tilde{G})$ as initial values, we then consider one-step expectation-maximization assisted estimators. For the k th stratum where $Z = z_k$, let y_{kl} , $l = 1, \dots, n_k$, denote the n_k distinct observed biomarker values and g_{kl} denote the corresponding point mass of $G_k(\cdot) = G(\cdot | z_k)$ such that $\sum_{l=1}^{n_k} g_{kl} = 1$. In the E-step, the conditional expectation of $I(Y = y_{kl})$ given $(X, \delta, Z = z_k, V = 0)$ and $\tilde{\theta}$ is

$$p_{kl}(\tilde{\theta}) = \frac{g_{kl} \lambda_{T(X|y_{kl}, \tilde{\theta})}^{\delta} S_{T(X|y_{kl}, \tilde{\theta})}}{\sum_{l=1}^{n_k} g_{kl} \lambda_{T(X|y_{kl}, \tilde{\theta})}^{\delta} S_{T(X|y_{kl}, \tilde{\theta})}}.$$

Therefore, for any function $h(\cdot)$ of Y , its conditional expectation is

$E\{h(Y) | \mathcal{O}, \tilde{\theta}\} = Vh(Y) + (1 - V) \sum_{k=1}^K I(Z = z_k) \sum_{l=1}^{n_k} h(y_{kl}) p_{kl}(\tilde{\theta})$. Let $\ell_n^F(\theta)$ denote the full likelihood function, where

$$\ell_n^F(\theta) = \sum_{i=1}^n [\delta_i \log \{\lambda_{T(X_i|Y_i)}\} - \log \{S_{T(X_i|Y_i)}\} + \log \{dG(Y_i|Z_i)\}].$$

In the M-step, we aim to maximize $\ell_n(\theta, \tilde{\theta}) = E\{\ell_n^F(\theta) | \mathcal{O}, \tilde{\theta}\}$, the expected full log-likelihood function given observed data and the current parameter estimates $\tilde{\theta}$ under the constraint $\sum_{l=1}^{n_k} g_{kl} = 1$, $k = 1, \dots, K$. Therefore, the one-step estimator of θ denoted by $\hat{\theta}^{(1)}$ could be obtained as the following,

$$\hat{\beta}^{(1)} = \beta: \sum_{i=1}^n \delta_i \left[E(Y_i | \mathcal{O}_i, \tilde{\theta}) - \frac{\sum_{j=1}^n E\{I(X_j \geq X_i) Y_j \exp(\beta Y_j) | \mathcal{O}_j, \tilde{\theta}\}}{\sum_{j=1}^n E\{I(X_j \geq X_i) \exp(\beta Y_j) | \mathcal{O}_j, \tilde{\theta}\}} \right] = 0,$$

$$\hat{\Lambda}^{(1)}(t) = \sum_{i=1}^n \frac{I(X_i \leq t) \delta_i}{\sum_{j=1}^n E\{I(X_j \geq X_i) \exp(\hat{\beta}^{(1)} Y_j) | \mathcal{O}_j, \tilde{\theta}\}},$$

$$\hat{G}_k^{(1)}(c) = n_k^{-1} \sum_{i=1}^n E\{I(Y_i \leq c, Z_i = z_k) | \mathcal{O}_i, \tilde{\theta}\}.$$

Plugging $\hat{\theta}^{(1)}$ in the estimating steps for the accuracy summaries will lead to the corresponding one-step estimators of $\widehat{\text{TPR}}_t^{(1)}(c)$, $\widehat{\text{FPR}}_t^{(1)}(c)$, $\widehat{\text{PPV}}_t^{(1)}(c)$ and $\widehat{\text{NPV}}_t^{(1)}(c)$, respectively.

3.3 Best Linear Unbiased Estimators

We show that the one-step estimators are consistent in the next section, however in general it may not be the case that one-step estimators are always more efficient compared with IPW estimator. To guarantee improvement on efficiency over IPW estimators, we propose to consider parameter-wise and element-wise best linear unbiased estimators (BLUE) combining IPW and one-step estimators with “best” meaning minimum variances. The parameter-wise BLUEs, $\hat{\theta}^{pBLUE}$ best linearly combine one-step and IPW estimators of θ_0 and is shown to be asymptotically equivalent to NPMLE in the next section with the corresponding weights for $\hat{\theta}^{(1)}$ and $\tilde{\theta}$ being functions of their influence functions. Therefore, NPMLEs could be identified without further iterations of EM algorithm if the best linear weights could be obtained. The corresponding accuracy measures could be estimated as plug-in estimators. However, the computation of the optimal weights is difficult due to the high dimension of the parameter space even though its analytical form is readily available. An element-wise BLUE (eBLUE) optimally and linearly combines one-step and IPW estimators of each univariate quantity of interests including β_0 , $\Lambda(t)$, $G_k(c)$ and any accuracy measures given c and fixed time t .

Specifically, we can construct an eBLUE for a generic accuracy measures $A_t(c)$ as $\hat{A}_t^{eBLUE}(c) = w_1 \hat{A}_t^{(1)}(c) + w_2 \tilde{A}_t(c)$ where w_1 and w_2 are weights that minimizes variance of $\hat{A}_t^{eBLUE}(c)$ with restriction $w_1 + w_2 = 1$. Let Υ denote the covariance matrix of $(\hat{A}_t^{(1)}, \tilde{A}_t(c))$, With simple derivation and as was shown in many existing work (e.g. Keller and Olkin, 2004), the best weights for $w = (w_1, w_2)$ is $(e^\top \Upsilon^{-1}) / (e^\top \Upsilon^{-1} e)$ with $e = (1, 1)^\top$ and

$$\text{var}(\hat{A}_t^{eBLUE}(c)) = 1 / (e^\top \Upsilon^{-1} e) < (e_2^\top \Upsilon^{-1} e_2) = \text{var}(\tilde{A}_t(c))$$

based on Cauchy-Schwartz inequality, where $e_2 = (0, 1)$.

4. ASYMPTOTIC RESULTS

4.1 Asymptotic Results for NPMLE of Accuracy Summaries

Let β_0 , Λ_0 , F_0 denote the true values of β , Λ and F . The asymptotic results of NPMLEs for regression parameters, cumulative baseline hazard function and covariate distribution under the semiparametric model for censored data have been extensively studied in Zeng et al. (2006) and Zeng and Lin (2014) for two-stage sampling design. Zeng and Lin (2014)

considered a more comprehensive setting where the completely observed covariates Z that are correlated with the biomarker of interest Y could be continuous and multidimensional. They showed the consistency of $\hat{\theta}$ and the weak convergence of $n^{1/2}(\hat{\beta} - \beta_0)$. Targeting accuracy measures of a single biomarker under two-stage sampling with a discrete surrogate variable Z , we will show both the consistency and the asymptotic normality of $\hat{\theta}$. The model considered in this paper is a special case of that in Zeng and Lin (2014) and thus the consistency of $\hat{\theta}$ follows the results in Zeng and Lin (2014). The weak convergence of $\hat{\theta}$ could be proved by applying appropriate regularity conditions. The technical details are provided in Appendix A and B of the Supplementary Materials.

We show that $n^{1/2}(\hat{\theta} - \theta_0) = n^{1/2}\mathcal{P}_n(\mathbb{W}_{\theta})$, where $\mathcal{P}_n$ denotes empirical measure and $\mathbb{W}_{\theta} = \{\mathcal{W}_{\beta}, \mathcal{W}_{\Lambda}, \mathcal{W}_G\}$ is given in the Appendix B. Using the results in Appendix B.3 of Liu et al. (2012), it follows that $n^{1/2}\{\hat{F}(c) - F(c)\} = n^{1/2}\mathcal{P}_n\{\mathcal{W}_F(c)\}$, $n^{1/2}\{\hat{S}_T(t|c) - S_T(t|c)\} = n^{1/2}\mathcal{P}_n\{\mathcal{W}_{S_T}(t|c)\}$, $n^{1/2}\{\hat{S}(t, c) - S(t, c)\} = n^{1/2}\mathcal{P}_n\{\mathcal{W}_S(t, c)\}$ where

$$\mathcal{W}_F(c) = \sum_{k=1}^K \left[p_k \mathcal{W}_{G_k}(c) + \{I(Z = z_k) - p_k\} G_k(c) \right],$$

$$\mathcal{W}_{S_T}(t|c) = -S_T(t) \exp(\beta_0 c) \{ \mathcal{W}_{\Lambda}(t) + c \Lambda(t) \mathcal{W}_{\beta} \},$$

$$\mathcal{W}_S(t, c) = \int_c^u \mathcal{W}_{S_T}(t|y) dF(y) + \int_c^u S_T(t|y) d\mathcal{W}_F(y).$$

Let $A_t(c)$ denote a generic term for accuracy measures evaluated at c for a given time t and $\hat{A}_t(c)$ be the corresponding NPMLE. Remembering that all the accuracy measures considered in this paper are simply the function of $F(c)$ and $S(t, c)$, we could thus show that $n^{1/2}\{\hat{A}_t(c) - A_t(c)\} = n^{1/2}\mathcal{P}_n\{\mathcal{W}_{A_t}(c)\}$, where $\mathcal{W}_{A_t}(c)$ are functions of $\mathcal{W}_F(c)$ and $\mathcal{W}_S(t, c)$ and are defined in the Supplementary Materials for each accuracy measures. Theorem 1 below summarizes the asymptotic results of $\hat{A}_t(c)$.

Theorem 1—Under conditions (a)–(d) in Appendix A, $\sup_{t,c} |\hat{A}_t(c) - A_t(c)| \rightarrow 0$ almost surely. In addition, $n^{1/2}\{\hat{A}_t(c) - A_t(c)\}$ converges weakly to a Gaussian process with mean zero and variance function $\Sigma_{A_t}(c) = \text{var}\{\mathcal{W}_{A_t}(c)\}$.

The uniform consistency of $\hat{A}_t(c)$ to $A_t(c)$ directly follows from the uniform consistency of $\hat{\theta}$ to θ_0 , and thus $\hat{F}(c)$ to $F(c)$ and $\hat{S}(t, c)$ to $S(t, c)$, along with a continuous mapping theorem

and lemma A.3 of Biliias et al. (1997). The weak convergence follows from the standard empirical process theory (van der Vaart and Wellner, 1996). The variance could be estimated by $\mathcal{P}_n \left\{ \mathcal{W}_{A_t}^2(c) \right\}$, where form of $\mathcal{W}_{A_t}(c)$ for each $A_t(c)$ is given in Appendix D.

4.2 Asymptotic Results for one-step Expectation-Maximization Assisted Estimators of Accuracy Summaries

Theorem 2 below summarizes the asymptotic results of one-step estimator $\hat{\theta}^{(1)}$.

Theorem 2—Under conditions (a)–(d), with probability one,

$$|\hat{\beta}^{(1)} - \beta_0| + \sup_{t \in [0, \tau]} |\hat{\Lambda}^{(1)}(t) - \Lambda_0(t)| + \sum_{k=1}^K \sup_{c \in [c_l, c_u]} |\hat{G}_k^{(1)}(c) - G_{0k}(c)| \rightarrow 0$$

almost surely. In addition, $n^{1/2}(\hat{\beta}^{(1)} - \beta_0, \hat{\Lambda}^{(1)} - \Lambda_0, \hat{G}^{(1)} - G_0)$ weakly converge to a zero-mean Gaussian process in $R \times \mathcal{L}^2[0, \tau] \times \mathcal{L}^2[c_l, c_u]^K$, where $\mathcal{L}^2[0, \tau]$ and $\mathcal{L}^2[c_l, c_u]$ are normed spaces consisting of all bounded functions and the norm is defined as the supremum norm on $\mathcal{B}_C$ and $\mathcal{B}_y$ respectively.

Justification for Theorem 2 was provided in Appendix C of supplementary materials.

Theorem 3 below summarizes the asymptotic results of $\hat{A}_t^{(1)}(c)$.

Theorem 3— $\sup_{t, c} |\hat{A}_t^{(1)}(c) - A_t(c)| \rightarrow 0$ almost surely. In addition, $n^{1/2} \left\{ \hat{A}_t^{(1)}(c) - A_t(c) \right\}$

converges weakly to a Gaussian process with mean zero and variance function

$$\Sigma_{A_t}^{(1)}(c) = \text{var} \left\{ \mathcal{W}_{A_t}^{(1)}(c) \right\}.$$

Similar to the proof of Theorem 1, the uniform consistency of $\hat{A}_t^{(1)}(c)$ to $A_t^{(t)}(c)$ rectly follows from the uniform consistency of $\hat{\theta}^{(1)}$ to θ_0 , and thus $\hat{F}^{(1)}(c)$ to $F(c)$ and $\hat{S}^{(1)}(t, c)$ to $S(t, c)$, along with a continuous mapping theorem and lemma A.3 of Biliias et al. (1997). The weak convergence of $\hat{A}_t^{(1)}(c)$ follows from the standard empirical process theory (van der Vaart and Wellner, 1996). Due to the iterative procedures of obtaining one-step EM estimators, the estimation of their corresponding asymptotic covariance is not straightforward even though the analytical form is available. We suggest that bootstrap sampling method be used for the estimation of covariances.

4.3 Asymptotic Results for Best Linear Unbiased Estimators of Accuracy Summaries

We provide analytical form of the optimal weights for the parameter-wise BLUEs of θ_0 in Appendix D of the supplementary material. The weak convergence of the pBLUE follows from asymptotic equivalence of $\hat{\theta}^{pBLUE}$ and $\hat{\theta}$, which is summarized in Theorem 4 below.

Theorem 4—Under conditions (a)–(d), with probability one,

$$\begin{aligned}
& |\hat{\beta}^{BLUE} - \beta_0| + \sup_{t \in [0, \tau]} |\hat{\Lambda}^{BLUE}(t) - \Lambda_0(t)| + \sum_{k=1}^K \sup_{c \in [c_l, c_u]} |\hat{G}_k^{BLUE}(c) - G_{0k}(c)| \xrightarrow{\mathcal{D}} |\hat{\beta} - \beta_0| \\
& + \sup_{t \in [0, \tau]} |\hat{\Lambda}(t) - \Lambda_0(t)| + \sum_{k=1}^K \sup_{c \in [c_l, c_u]} |\hat{G}_k(c) - G_{0k}(c)|
\end{aligned}$$

The proof of Theorem 4 is provided in Appendix D of the supplementary material. The weak convergence of pBLUEs of accuracy measures could be obtained similarly as the proof of Theorem 3 for one-step estimators of accuracy measures. Although the analytical form of optimal weights for pBLUE is available, the computation is not straightforward due to the high dimension of the parameter space. We suggest to consider element-wise BLUEs for computational feasibility.

The weak convergence of $\hat{A}_t^{eBLUE}(c)$ is summarized in Theorem 5 below.

Theorem 5— $\sup_{t,c} |\hat{A}_t^{eBLUE}(c) - A_t(c)| \rightarrow 0$ almost surely. In addition, $n^{1/2} \left\{ \hat{A}_t^{eBLUE}(c) - A_t(c) \right\}$ converges weakly to a Gaussian process with mean zero and variance function $\Sigma_{A_t}^{eBLUE}(c) = (e^\top \Upsilon e)^{-1}$, where $\Upsilon = \text{cov} \left\{ \mathcal{W}_{A_t}^{(1)}(c), \mathcal{W}_{A_t}^{IPW}(c) \right\}$.

The proof of Theorem 5 directly follows from the proof of weak convergence of the one-step estimator (Theorem 3) and IPW estimator as well as results in Section 3.3. The expression of $\mathcal{W}_{A_t}^{(1)}(c)$ is provided in Appendix C.2 of the supplementary material and the expression of $\mathcal{W}_{A_t}^{IPW}(c)$ is provided in Liu et al. (2012). We suggest the bootstrap sampling method be used for the estimation of Υ .

In the next section, we compare their finite sample performance with IPW, one-step, eBLUE and NPMLE estimators using numerical study.

5. ESTIMATION INCREMENTAL VALUES OF A NEW MARKER

The approach for evaluating a single marker can be extended to the estimation of the improvement of the aforementioned accuracy summaries by a novel marker beyond an existing prognostic score. We now define $\mathbf{Y}$ as the vector of all potential risk predictors, $\mathbf{Y}_{oldi}$ includes a set of routine markers available for all, Y_{newi} represents a novel risk marker only ascertained at the second phase for selected a subset of individuals, and Z_i represents auxiliary variables including matching and stratification variables available for all members in the cohort and are under consideration for matching or stratification. While both Z_i and $\mathbf{Y}_{oldi}$ are available for the entire cohort, Y_{newi} is only available if $V_i = 1$, where V_i is a binary variable indicating whether subject i is selected to the phase II subcohort. Let $\mathcal{R}_t(\mathbf{Y})$ denote the risk score $P(T_i \leq t | \mathbf{Y}_i)$. The accuracy summaries is now defined as

$$\text{TPR}_t(c) = \text{pr}(\mathcal{R}_t(\mathbf{Y}) > c | T \leq t), \text{ and } \text{FPR}_t(c) = \text{pr}(\mathcal{R}_t(\mathbf{Y}) > c | T > t).$$

Similarly to the single marker evaluation specified in Section 2, we again assume that the risk $\mathcal{R}_t(\mathbf{Y}) = P(T_i \leq t | \mathbf{Y}_i)$ follows a semi-parametric transformation model (Cheng et al., 1995, 1997; Zeng and Lin, 2006)

$$1 - \mathcal{R}_t(\mathbf{Y}) = \mathcal{F}_0 \left[\log \{H(t)\} + \beta_{\text{new}} Y_{\text{new}i} + \beta_{\text{old}}^\top \mathbf{Y}_{\text{old}i} \right],$$

The likelihood specified in Equation 2 can be modified as

$$\mathcal{L}_n(\theta) = \prod_{i=1}^n \left\{ \lambda_T(X_i | \mathbf{Y}_i)^{\delta_i} S_T(X_i | \mathbf{Y}_i) dG(\mathbf{Y}_i | Z_i) \right\}^{V_i} \times \left\{ \int_{c_l}^{c_u} \lambda_T(X_i | c)^{\delta_i} S_T(X_i | c) dG(c | Z_i) \right\}^{1 - V_i}.$$

Since all components in $\mathbf{Y}_{\text{old}}$ are continuous, to estimate $G(\mathbf{Y}_i | Z_i)$ one may consider first estimate $G(Y_{\text{new}i} | \mathbf{Y}_{\text{old}i}, Z_i)$ using a semiparametric regression quantile method (Heagerty and

Pepe, 1999). Specifically, we let $Y_{\text{new}i} = \mu(\mathbf{Y}_{\text{old}i}, Z_i) + \sigma(\mathbf{Y}_{\text{old}i}, Z_i)\varepsilon(\mathbf{Y}_{\text{old}i}, Z_i)$, where $\mu(\mathbf{Y}_{\text{old}i}, Z_i)$, $\sigma^2(\mathbf{Y}_{\text{old}i}, Z_i)$ and $\sigma(\mathbf{Y}_{\text{old}i}, Z_i)$ are respectively the mean, variance and standardized residual of $Y_{\text{new}i}$. Flexible kernel based procedure can be used to estimate $G(\varepsilon |$

$\mathbf{Y}_{\text{old}i}, Z_i)$ as $\hat{G}(\varepsilon) = \sum_i^n \hat{\pi}_i(\varepsilon, h) I(\varepsilon_i \leq \varepsilon)$ with $\hat{\pi}(\varepsilon, h) = \frac{\sum_{i=1}^n V_i K_h(\varepsilon - \varepsilon_i)}{\sum_{i=1}^n K_h(\varepsilon - \varepsilon_i)}$ where $K_h(\cdot) = K(\cdot/h)$

is a symmetric kernel density function with bandwidth h . Similar approach to deal with continuous covariates has also been considered in Zeng and Lin (2014). The theoretical properties of incremental values of accuracy summaries of $\mathcal{R}_t(\mathbf{Y})$ over that of $\mathcal{R}_t(\mathbf{Y}_{\text{old}})$ are similar to that developed in the semiparametric situation (Zhou et al., 2015).

6. SIMULATION

Simulation studies were conducted to assess the finite sample properties of the proposed method. Failure time T was generated from Cox model $\lambda_T(t | Y) = \lambda(t) \exp(\beta Y)$, with $\lambda(t) = 1$, $\beta = \log(3)$. The continuous biomarker Y was generated from truncated normal such that $Y = \text{sign}(\tilde{Y}) \min(|\tilde{Y}|, 5)$ with $\tilde{Y} \sim N(0, 1)$. The censoring time $C = \min(\tilde{C}, 2)$ with $\tilde{C} \sim \text{Gamma}(2.5, 2)$, resulting in about 84% of censoring.

The first simulation study considered the stratified case-cohort design with full cohort of size $n = 2000$. The stratum variable defined by the combination of a surrogate variable $Z = \mathbb{I}(Y > 0)$ and the case/noncase status had four levels: case with $Y > 0$ (3%), case with $Y = 0$ (12%), control with $Y > 0$ (47%), and control with $Y = 0$ (38%). The subcohort sample size was 200 with 50 subjects were randomly selected from each strata, resulting in a ratio of 1:1 between individuals with and without an event. Results based on 1000 simulated datasets were summarized for each measure of Y in predicting the risk of having an event by time $t_0 = 1$ in Table 1. The unbiasedness and efficiency were compared among the proposed one-

step estimators, element-wise best linear unbiased estimators (eBLUE), the proposed NPMLEs as well as the IPW estimators. We presented results for $\text{TPR}_{\lambda(c)}$, $\text{FPR}_{\lambda(c)}$, $\text{PPV}_{\lambda(c)}$ and $\text{NPV}_{\lambda(c)}$ at threshold values taken to be the 25th, 50th and 75th percentile of Y , i.e., $c = \mathcal{F}^{-1}(p)$ for $p = 0.25, 0.5, 0.75$. (Table 1a), and results for global measures AUC and DMR, as well as FPR, PPV and NPV at a TPR level of 0.9 (Table 1b). It appeared that the one-step, eBLUE estimators and NPMLEs were essentially unbiased and the average of the estimated analytical standard errors (ASE) are similar to their empirical standard deviations (ESD) with coverage probabilities (CP) near the nominal level of 95%. Worth mentioning, all three types of estimators substantially improved the efficiency of the simple IPW estimators, and for many accuracy summaries the efficiency gains are quite comparable among the three types of proposed estimators. Nevertheless the NPMLEs are more computationally intensive as their averaged and maximum number of iterations is 65 and 88 respectively as compared to one iteration for the one-step estimators and eBLUEs.

The second simulation study considered the matched nested case-control design with full cohort of size $N = 1000$. One control per case was selected with the matching variable again being defined by $Z = \mathbb{I}(Y > 0)$. Results based on 1000 simulated datasets were summarized in Table 2. We compared the empirical performance of the proposed one-step, the eBLUE estimators, and the NPMLEs as well as the IPW estimators in the same way as the first simulation study. Similar to the results observed with the stratified case-cohort design, all three types of proposed estimators provided evident improvement over the IPW estimator. The averaged and maximum number of iteration for the NPMLEs is 60 and 135. We observed that with the simulated sample sizes the empirical coverages for the eBLUE estimators are slightly lower than the nominal level in Table 1 and Table 2. The coverages improved when larger sample sizes were considered (data not shown).

Our proposed estimators require that censoring time C to be independent of Y and Z . We conducted simulation study to investigate how the estimators perform under a situation when C is dependent on Y . Specifically, we repeated the first two sets of simulation studies but with C generated as $C = \min(\exp(0.1Y), \tilde{C})$ with $\tilde{C} \sim \text{Gamma}(2, 2)$. In addition, the matching/stratum variable Z was generated as $Z = \mathbb{I}(Y^* > 0)$, where Y^* was generated from a standard normal distribution and had a correlation of 0.5 with Y . We found that under a sCCH design, the three types of proposed estimators were only slightly biased, with percentage of bias all below 5% (see Table 3). With a mNCC design, the one-step and eBLUE estimators remained only slightly biased, with bias percentages all below 4%. NPMLEs seemed to be more biased, with percentage of bias reached as high as 9.89% (Table 4). Under both designs, as expected, IPW estimators remained unbiased. It appeared that the one-step and eBLUE estimators can be quite robust under the marker dependent censoring, this might be due to both the facts that the matching variable Z was correlated with Y , and also that the one-step and eBLUE estimators were both more directly derived from the unbiased IPW estimators. Note also that all three proposed estimators still had smaller MSE compared with that of IPW estimators under such a misspecified setting.

7. APPLICATION

Detection of patients with hepatocellular carcinoma (HCC) at early stage when curative interventions can be implemented can lead to good prognosis. Surveillance of high-risk populations with cirrhosis or chronic infection with hepatitis B virus (HBV) or hepatitis C virus (HCV) is therefore important in detecting early HCC. Alpha fetoprotein (AFP) is the most widely used biomarker for HCC surveillance, however its sensitivity and specificity in detecting early HCC are low. More reliable biomarkers for HCC surveillance and early detection are sought in order to improve the outcome of the disease.

The Hepatitis C Antiviral Long-Term Treatment against Cirrhosis (HALT-C) Trial enrolled 1050 patients with chronic hepatitis C and bridging fibrosis or cirrhosis and failure to achieve sustained virologic response to pegylated interferon and ribavirin therapy, who were followed every 3 months during the 3.5 years of the randomized trial of low dose pegylated interferon vs. no treatment. Blood samples were collected at each visit for subsequent research testing including assays for HCC biomarkers. Ultrasound examinations were at enrollment and repeated 6 months after enrollment and again every 12 months. Patients with elevated or rising AFP and those with new lesions on ultrasound were evaluated further by CT or MRI.

One of the goals of the HALT-C Trial was to identify and validate markers for the surveillance of HCC, NS in particular, to evaluate the accuracy performance of a novel serum biomarker, des- γ -carboxy prothrombin (DCP), for predicting the diagnosis of HCC. A nested case-control study was used to evaluate the accuracy of DCP in the detection of HCC at times before the diagnosis of HCC. For this study, all 39 HCC cases diagnosed between randomization and 3.8 years after randomization were included. For each case, 2 controls without HCC, matched for treatment assignment, presence of cirrhosis on baseline biopsy and length of follow-up, were selected. Blood samples at the time of HCC diagnosis and selected followup times were tested for AFP and DCP. The results of the nested case-control study of repeated markers were reported in Lok et al. (2010), where conditional logistic regression models were used to compare characteristics of HCC cases and matched controls, and unconditional logistic regression was used to evaluate the accuracy performance of the biomarkers.

For illustration of our proposed method, only baseline measurements of the 39 cases and 77 matched controls selected in the HALT-C nested case-control study are considered. Due to the low number of liver cancer cases in the nested case-control study, methods that attempt to exploit study information available as much as possible in order to improve efficiency is particularly appealing. Logarithm transformed values were considered for DCP (denote by $\log(\text{DCP})$). In a Cox regression model using $\log(\text{DCP})$ as a covariate, hazards ratio is 3.79 (95% CI: (1.55, 9.28)) for $\log(\text{DCP})$, suggesting that DCP values increased as time to HCC diagnosis approached. We subsequently formally evaluate the t -year predictive performance of $\log(\text{DCP})$ using the measures described in Section 3.1 based on the nested case-control subset. Since the nested case-control study was assembled 3.8 years after randomization, we set $t_0 = 3$ years to evaluate the predictiveness of biomarkers in predicting HCC 3 years after marker measurement. We used the proposed NPML estimators in order to account for

nested case-control sampling. In addition, we calculated the one-step and eBLUE estimators as well as the semiparametric IPW estimators for comparison. As seen from Table 5, log(DCP) alone can signify the onset of HCC as early as 3 years in advance of the diagnosis, with 3-year AUC equals 0.684 with 95% confidence interval (0.635, 0.733) based on the NPMLE estimator. Similar conclusion can be made with IPW estimators but with a wide interval estimates. The 3-year AUC is 0.722 with 95% confidence interval (0.628, 0.816). On almost all parameters of interest, we observed that NPMLE estimators can lead to substantial improvement in efficiency, and improvement can be observed in some cases with just iteration of our proposed hybrid estimators. We also found in this particular example the NPMLE estimators are quite more efficient than the hybrid estimators in a few cases, suggesting that it is beneficial to let NPMLE converge for additional gain in efficiency in this case. The example illustrated that the extra computational effort in the maximum likelihood based approach can be helpful in the situations with small number of cases. Such situations are quite common in cancer research due to low incidence rate or short follow-up of the research cohort.

8. REMARKS

Statistical methods aiming to improve efficiency for inference is important in analyzing two-phase biomarker studies of rare diseases such as cancer. In contrast to IPW estimators, the proposed NPMLE estimators for predictive performance assessment optimally use available information, and therefore can lead to substantial gain in efficiency. While more computational intensive than the IPW estimators, both our numerical studies and the example of clinical application indicate that it is still worthwhile considering the class of likelihood-based estimators as a valuable alternative to the IPW estimators to achieve more precise inferences.

The proposed hybrid estimators also lead to substantial gain in efficiency compared with the IPW estimators, while alleviating the computational burden of the NPMLE estimators. In our simulations, the one-step or eBLUE estimators often achieve efficiency comparable to the NPMLE estimators for a majority of accuracy summaries. Such improvement would be very valuable in practice when there is greater uncertainty regarding the aptness of the assumed model and the convergence issue becomes salient. Compared with ordinary NPMLE estimator, the calculation of one-step or eBLUE estimators requires that consistent estimators under the two-phase studies are available. However, in practice, the sampling probability might either be difficult to calculate or biased due to practical issues such as inaccurate documentation of the original design or inappropriate execution of the sampling design. As such, the standard IPW estimators might not be available or biased. For such settings, one may use an approximately correct initial estimator and let the iterations fully converge to obtain the NPMLE, which remains valid regardless of the consistency of the initial estimator.

Both NPMLE and hybrid estimators require the assumption of independent censoring. We have investigated the robustness of the proposed estimators when censoring is related to Y . We found that under the numerical settings we considered the one-step and eBLUE estimators tended to be quite robust compared with the NPMLE estimators, especially when

the matching variables were correlated with Y . We have briefly described an extension to the settings when one needs to incorporate other markers available from the full cohort in the prognostic model, however our proposed procedures can be computationally infeasible when the dimension of biomarkers collected only at the second phase increases. Approaches that handle these complexities warrants future research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work is supported by grants U01-CA86368, P01-CA053996, R01-GM085047, R01-GM079330 and R01-ES016626 awarded by the National Institutes of Health.

References

- Andersen P, Gill R. Cox's regression model for counting processes: a large sample study. *The annals of statistics*. 1982; 10(4):1100–1120.
- Bilias Y, Gu M, Ying Z. Towards a general asymptotic theory for cox model with staggered entry. *The Annals of Statistics*. 1997; 25(2):662–682.
- Borgan O, Langholz B, Samuelsen S, Goldstein L, Pogoda J. Exposure stratified case-cohort designs. *Lifetime Data Analysis*. 2000; 6(1):39–58. [PubMed: 10763560]
- Breslow N, Wellner J. Weighted likelihood for semiparametric models and two-phase stratified samples, with application to cox regression. *Scand J Statist*. 2007; 34(1):86–102.
- Cai T, Zheng Y. Nonparametric evaluation of biomarker accuracy under nested case-control studies. *Journal of the American Statistical Association*. 2011; 106(494):569–580. [PubMed: 22844169]
- Cai T, Zheng Y. Evaluating prognostic accuracy of biomarkers in nested case-control studies. *Biostatistics*. 2012; 13(1):89–100. [PubMed: 21856652]
- Chen H. Double-semiparametric method for missing covariates in cox regression models. *Journal of the American Statistical Association*. 2002; 97(458):565–576.
- Chen K. Generalized case-cohort sampling. *Journal of the Royal Statistical Society, Series B, Methodological*. 2001; 63(4):791–809.
- Cheng S, Wei L, Ying Z. Analysis of transformation models with censored data. *Biometrika*. 1995; 82(4):835.
- Cheng S, Wei L, Ying Z. *Journal of the American Statistical Association*. 1997. Predicting survival probabilities with semiparametric transformation models; 227–235.
- Cook NR, Paynter NP, Eaton CB, Manson JE, Martin LW, Robinson JG, Rossouw JE, Wassertheil-Smoller S, Ridker PM. Comparison of the framingham and reynolds risk scores for global cardiovascular risk prediction in the multiethnic women's health initiative. *Circulation*. 2012; 125(14):1748–1756. [PubMed: 22399535]
- Cox D. Regression models and life-tables. *Journal of the Royal Statistical Society. Series B (Methodological)*. 1972:187–220.
- Folsom A, Aleksic N, Catellier D, Juneja H, Wu K. C-reactive protein and incident coronary heart disease in the atherosclerosis risk in communities (aric) study. *American Heart Journal*. 2002; 144(2):233–238. [PubMed: 12177639]
- Goldstein L, Langholz B. Asymptotic theory for nested case-control sampling in the Cox regression model. *Ann Statist*. 1992; 20:1903–28.
- Gray R. Weighted analyses for cohort sampling designs. *Lifetime Data Analysis*. 2009; 15(1):24–40. [PubMed: 18712477]

- Heagerty PJ, Pepe MS. Semiparametric estimation of regression quantiles with application to standardizing weight for height and age in us children. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 1999; 48(4):533–551.
- Keller T, Olkin I. Combining correlated unbiased estimators of the mean of a normal distribution. 2004:218–227. *Lecture Notes-Monograph Series*
- Langholz B, Borgan Y. Estimation of absolute risk from nested case-control data. *Biometrics*. 1997; 53:767–774. [PubMed: 9192463]
- Lin D. On fitting Cox's proportional hazards models to survey data. *Biometrika*. 2000; 87(1):37.
- Liu D, Cai T, Zheng Y. Evaluating the predictive value of biomarkers with stratified case-cohort design. *Biometrics*. 2012; 68(4):1219–1227. [PubMed: 23173848]
- Lok AS, Sterling RK, Everhart JE, Wright EC, Hoefs JC, Di Bisceglie AM, Morgan TR, Kim HY, Lee WM, Bonkovsky HL, et al. Des- γ -carboxy prothrombin and α -fetoprotein as biomarkers for the early detection of hepatocellular carcinoma. *Gastroenterology*. 2010; 138(2):493–502. [PubMed: 19852963]
- Pollard D. *Empirical processes: theory and applications*. Institute of Mathematical Statistics; 1990.
- Prentice R. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika*. 1986; 73(1):1.
- Prentice R, Breslow N. Retrospective studies and failure time models. *Biometrika*. 1978; 65(1):153–158.
- Robinson PM. The Stochastic Difference between Econometric Statistics. *Econometrica*. 1988; 56(3): 531–48.
- Samuelson SO. A pseudolikelihood approach to analysis of nested case-control studies. *Biometrika*. 1997; 84:379–394.
- Scheike T, Juul A. Maximum likelihood estimation for cox's regression model under nested case-control sampling. *Biostatistics*. 2004; 5(2):193–206. [PubMed: 15054025]
- Scheike T, Martinussen T. Maximum likelihood estimation for Cox's regression model under case-cohort sampling. *Scandinavian journal of statistics*. 2004; 31(2):283–293.
- Thomas DC. Addendum to "Methods of cohort analysis: Appraisal by application to asbestos mining". *Journal of the Royal Statistical Society, Series A, General*. 1977; 140:483–485.
- van der Vaart AW, Wellner JA. *Weak Convergence and Empirical Process: With Applications to Statistics*. Springer; 1996.
- Wang T, Rohan T, Gunter M, Xue X, Wactawski-Wende J, Rajpathak S, Cushman M, Strickler H, Kaplan R, Wassertheil-Smoller S, Scherer P, GY H. A prospective study of inflammation markers and endometrial cancer risk in postmenopausal hormone nonusers. *Cancer Epidemiology Biomarkers & Prevention*. 2011; 20(5):971–7.
- Zeng D, Lin D. Maximum likelihood estimation in semiparametric transformation models for counting processes. *Biometrika*. 2006; 93:627–40.
- Zeng D, Lin D. Efficient estimation of semiparametric transformation models for two-phase cohort studies. *Journal of the American Statistical Association*. 2014; 0(ja) null.
- Zeng D, Lin D, Avery C, North K, Bray M. Efficient semiparametric estimation of haplotype-disease associations in case-cohort and nested case-control studies. *Biostatistics*. 2006; 7(3):486–502. [PubMed: 16500923]
- Zhou QM, Zheng Y, Chibnik LB, Karlson EW, Cai T. Assessing incremental value of biomarkers with multi-phase nested case-control studies. *Biometrics*. 2015; 71(4):1139–1149. [PubMed: 26195245]

Table 1

Case cohort design with a 2-level stratum variable: Finite sample performance of the proposed one-step EM estimators, eBLUE estimators and NPMLEs as well as the IPW with full cohort size $N = 2000$ and subcohort size $N_1 = N_2 = 50$ for events, $N_1 = N_2 = 50$ for nonevents. Shown below are percent in bias (PBias), empirical standard deviation (ESD) and average of the estimated standard error (ASE) of the proposed estimators, and empirical coverage level of the 95% confidence intervals (CP) as well as mean square error (MSE).

(a) Accuracy estimates at $c_p = \mathcal{F}^{-1}(p)$ for $p = 0.25, 0.5, 0.75$.

	one-step						eBLUE						NPMLE						IPW					
	True	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP	MSE	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE
$TPR_{t_0}^{(c, .25)}$	0.953	0.08	0.012	0.012	92.1	0.014	0.19	0.012	0.011	90.6	0.014	-0.00	0.010	0.010	94.8	0.010	0.08	0.020	0.08	0.020				
$TPR_{t_0}^{(c, .5)}$	0.834	0.17	0.024	0.023	93.4	0.056	0.25	0.023	0.023	92.6	0.055	-0.02	0.020	0.019	94.3	0.040	0.22	0.091	0.22	0.091				
$TPR_{t_0}^{(c, .75)}$	0.602	0.13	0.045	0.044	93.5	0.205	0.08	0.045	0.043	93.3	0.199	-0.22	0.041	0.040	93.2	0.171	0.35	0.271	0.35	0.271				
$FPR_{t_0}^{(c, .25)}$	0.715	0.39	0.037	0.036	94.3	0.134	0.41	0.036	0.036	94.2	0.133	0.25	0.029	0.029	95.0	0.083	0.34	0.150	0.34	0.150				
$FPR_{t_0}^{(c, .5)}$	0.443	0.19	0.012	0.012	95.4	0.015	0.19	0.012	0.012	95.4	0.015	0.17	0.012	0.012	95.6	0.015	0.05	0.019	0.05	0.019				
$FPR_{t_0}^{(c, .75)}$	0.190	-0.06	0.027	0.027	93.6	0.072	-0.15	0.027	0.026	93.1	0.072	0.10	0.025	0.025	94.6	0.063	-0.23	0.078	-0.23	0.078				
$PPV_{t_0}^{(c, .25)}$	0.185	-0.66	0.014	0.014	95.3	0.019	-0.97	0.014	0.014	94.4	0.020	-0.21	0.012	0.012	94.6	0.014	0.01	0.037	0.01	0.037				
$PPV_{t_0}^{(c, .5)}$	0.243	-0.55	0.016	0.016	95.1	0.025	-0.80	0.016	0.016	94.8	0.026	-0.22	0.015	0.015	94.6	0.022	0.17	0.060	0.17	0.060				
$PPV_{t_0}^{(c, .75)}$	0.350	0.19	0.031	0.032	95.8	0.094	-0.08	0.031	0.031	94.9	0.096	0.18	0.029	0.029	95.2	0.086	0.95	0.190	0.95	0.190				
$NPV_{t_0}^{(c, .25)}$	0.973	0.03	0.007	0.006	91.9	0.004	0.08	0.006	0.006	91.3	0.004	-0.01	0.005	0.005	93.6	0.003	0.02	0.006	0.02	0.006				

(a) Accuracy estimates at $c_p = \mathcal{F}^{-1}_p(p)$ for $p = 0.25, 0.5, 0.75$.

	$\hat{\tau}$																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																														
	one-step												eBLUE						NPMLE						IPW																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																						
	True	PBias	ESD	ASE	CP	MSE	 

(b) Accuracy estimates at $\text{TPR}_{t_0} = 0.9, \text{AUC}_{t_0}$ and DMR_{t_0}

	one-step						eBLUE						NPMLE						IPW	
	True	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP	MSE	PBias	MSE		
$FPR_{t_0}^*$	0.568	-1.19	0.041	0.040	93.4	0.171	-1.05	0.038	0.036	93.4	0.145	-0.38	0.034	0.036	96.4	0.119	-1.73	0.298		
$PPV_{t_0}^*$	0.212	0.49	0.016	0.017	95.5	0.027	0.09	0.016	0.017	95.1	0.027	0.34	0.016	0.015	94.8	0.025	1.80	0.070		
$NPV_{t_0}^*$	0.962	-0.03	0.005	0.005	94.8	0.003	-0.01	0.005	0.005	94.2	0.003	-0.04	0.004	0.004	93.3	0.002	-0.06	0.004		
AUC_{t_0}	0.786	-0.58	0.020	0.022	96.0	0.043	-0.57	0.018	0.019	95.3	0.035	-0.81	0.017	0.020	96.6	0.034	-0.48	0.074		
DMR_{t_0}	0.292	-1.11	0.020	0.020	94.8	0.042	-1.01	0.018	0.018	93.3	0.035	-1.66	0.018	0.017	93.9	0.034	-0.86	0.075		

* evaluated at $\text{TPR}_{t_0} = 0.9$.

** MSE is in unit of 1e-2.

Table 2

Nested case control design with a two-level matching variable: Finite sample performance of the proposed one-step EM estimators, eBLUE estimators and NPMLEs as well as the IPW with full cohort size $N = 1000$ and 1 matched control. Shown below are percent in bias (PBias), empirical standard deviation (ESD) and average of the estimated standard error (ASE) of the proposed estimators, and empirical coverage level of the 95% confidence intervals (CP) as well as mean square error (MSE).

(a) Accuracy estimates at $c_p = \mathcal{F}^{-1}(p)$ for $p = 0.25, 0.5, 0.75$.

	one-step												eBLUE						NPMLE						IPW				
	True	PBias			ESD			ASE			CP			MSE			PBias	ESD			ASE			CP			MSE	PBias	MSE
		True	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP		MSE	PBias	ESD	ASE	CP	MSE	PBias	MSE				
$TPR_{t_0}^{(c;.25)}$	0.953	-0.03	0.016	0.015	90.1	0.026	0.24	0.016	0.014	86.6	0.026	0.01	0.012	0.012	93.7	0.014	-0.01	0.034											
$TPR_{t_0}^{(c;.5)}$	0.834	-0.02	0.025	0.025	94.4	0.062	0.25	0.025	0.025	92.4	0.065	-0.07	0.023	0.023	95.5	0.055	0.01	0.120											
$TPR_{t_0}^{(c;.75)}$	0.602	0.05	0.038	0.037	93.3	0.145	0.18	0.038	0.037	93.1	0.146	0.01	0.036	0.036	94.2	0.133	0.05	0.208											
$FPR_{t_0}^{(c;.25)}$	0.715	0.51	0.062	0.058	90.8	0.387	1.18	0.062	0.056	90.0	0.391	0.44	0.042	0.041	94.3	0.178	0.58	0.476											
$FPR_{t_0}^{(c;.5)}$	0.443	0.20	0.018	0.016	92.0	0.031	0.20	0.018	0.016	92.2	0.031	0.16	0.018	0.017	93.4	0.031	0.48	0.189											
$FPR_{t_0}^{(c;.75)}$	0.190	0.69	0.021	0.021	93.3	0.045	0.24	0.021	0.021	93.3	0.046	0.62	0.019	0.019	94.3	0.037	0.90	0.087											
$PPV_{t_0}^{(c;.25)}$	0.185	-0.73	0.018	0.019	94.4	0.034	-1.27	0.019	0.017	91.8	0.038	-0.50	0.016	0.017	94.8	0.026	0.14	0.042											
$PPV_{t_0}^{(c;.5)}$	0.243	-0.82	0.020	0.021	95.0	0.040	-0.98	0.020	0.021	95.2	0.039	-0.48	0.019	0.020	95.2	0.036	0.00	0.046											
$PPV_{t_0}^{(c;.75)}$	0.350	-0.71	0.033	0.033	93.2	0.110	-0.67	0.033	0.033	93.1	0.108	-0.47	0.031	0.031	94.0	0.098	0.08	0.118											
$NPV_{t_0}^{(c;.25)}$	0.973	-0.00	0.007	0.007	93.3	0.005	0.02	0.007	0.006	90.1	0.004	-0.01	0.006	0.006	94.4	0.004	-0.01	0.005											

(a) Accuracy estimates at $c_p = \mathcal{F}^{-1}_p(p)$ for $p = 0.25, 0.5, 0.75$.

	τ																										
	one-step												eBLUE						NPMLE						IPW		
	True	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP	MSE	PBias
$NPV_{t_0}^{(c_{.5})}$	0.952	0.02	0.008	0.009	94.8	0.007	0.06	0.008	0.007	89.6	0.007	-0.01	0.008	0.008	95.3	0.006	0.00	0.008	0.008	95.3	0.006	0.00	0.008	0.008	0.00	0.008	0.008
$NPV_{t_0}^{(c_{.75})}$	0.923	0.06	0.010	0.010	93.9	0.009	0.12	0.010	0.008	88.6	0.010	0.02	0.009	0.009	94.5	0.009	-0.01	0.011	0.009	94.5	0.009	-0.01	0.011	0.009	0.00	0.011	0.011

(b) Accuracy estimates at $\text{TPR}_{t_0} = 0.9, \text{AUC}_{t_0}$ and DMR_{t_0}

	one-step						eBLUE						NPMLE						IPW	
	True	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP	MSE	PBias	ESD	ASE	CP	MSE	PBias	MSE		
$FPR^*_{t_0}$	0.568	-1.93	0.046	0.045	92.6	0.226	-1.89	0.046	0.044	92.1	0.222	-0.58	0.040	0.043	95.6	0.165	-2.43	0.280		
$PPV^*_{t_0}$	0.212	0.69	0.021	0.021	94.8	0.042	0.31	0.021	0.021	94.2	0.044	0.23	0.019	0.020	94.9	0.038	1.87	0.059		
$NPV^*_{t_0}$	0.962	-0.12	0.006	0.006	97.2	0.003	-0.09	0.006	0.006	95.2	0.003	-0.06	0.005	0.005	94.6	0.002	-0.20	0.004		
AUC_{t_0}	0.786	-1.79	0.021	0.028	96.0	0.065	-1.87	0.021	0.027	95.5	0.066	-2.10	0.017	0.026	95.9	0.057	-1.68	0.069		
DMR_{t_0}	0.292	-0.41	0.021	0.021	94.7	0.046	-0.39	0.021	0.021	95.0	0.044	-1.19	0.020	0.019	94.5	0.041	-0.35	0.055		

* evaluated at $\text{TPR}_{t_0} = 0.9$.

** MSE is in unit of 1e-2.

sCCH Under dependent censoring: Finite sample performance of the proposed one-step EM estimators, eBLUE estimators and NPMLEs as well as the IPW with full cohort size $N = 2000$ and subcohort size $N_1 = N_2 = 50$, $N_3 = N_4 = 50$. Shown below are percent in bias (PBias) and mean square error (MSE).

Table 3

(a) Accuracy estimates at $c_p = \mathcal{F}^{-1}(p)$ for $p = 0.25, 0.5, 0.75$.

	one-step			eBLUE			NPMLE			IPW		
	True	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE	MSE
$TPR_{t_0}^{(c, .25)}$	0.953	-0.00	0.020	0.17	0.019	0.16	0.015	0.00	0.025			
$TPR_{t_0}^{(c, .5)}$	0.834	0.15	0.100	0.43	0.093	0.70	0.076	0.16	0.123			
$TPR_{t_0}^{(c, .75)}$	0.602	0.53	0.228	0.76	0.211	1.78	0.178	0.51	0.278			
$FPR_{t_0}^{(c, .25)}$	0.715	0.00	0.173	0.07	0.172	0.65	0.146	-0.12	0.180			
$FPR_{t_0}^{(c, .5)}$	0.443	-0.08	0.186	-0.00	0.186	1.63	0.159	-0.41	0.198			
$FPR_{t_0}^{(c, .75)}$	0.190	0.08	0.107	0.07	0.107	3.29	0.094	-0.66	0.114			
$PPV_{t_0}^{(c, .25)}$	0.185	2.15	0.023	2.07	0.024	2.86	0.022	0.41	0.037			
$PPV_{t_0}^{(c, .5)}$	0.243	2.40	0.050	2.30	0.050	2.41	0.046	0.96	0.077			
$PPV_{t_0}^{(c, .75)}$	0.350	2.77	0.164	2.77	0.165	2.19	0.143	1.78	0.228			
$NPV_{t_0}^{(c, .25)}$	0.973	-0.07	0.006	-0.02	0.006	-0.05	0.004	0.02	0.007			
$NPV_{t_0}^{(c, .5)}$	0.952	-0.08	0.009	-0.04	0.009	-0.07	0.006	0.05	0.010			

(a) Accuracy estimates at $c_p = \mathcal{F}^{-1}(\rho)$ for $p = 0.25, 0.5, 0.75$.

$NPV_{t_0}(c_{.75})$	one-step						eBLUE			NPMLE			IPW		
	True	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE
	0.923	-0.13	0.011	-0.09	0.011	-0.13	0.008	-0.13	0.008	0.07	0.013	0.07	0.013		

(b) Accuracy estimates at $TPR_{t_0} = 0.9, AUC_{t_0}$ and DMR_{t_0}

	one-step			eBLUE		NPMLE		IPW	
	True	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE
$FPR_{t_0}^*$	0.568	-1.51	0.261	-1.54	0.245	-1.27	0.207	-1.97	0.340
$ppv_{t_0}^*$	0.212	3.39	0.043	3.41	0.043	4.19	0.045	2.15	0.075
$NPV_{t_0}^*$	0.962	-0.15	0.004	-0.13	0.004	-0.18	0.003	-0.05	0.004
AUC_{t_0}	0.786	-0.49	0.068	-0.32	0.062	-0.58	0.056	-0.36	0.086
DMR_{t_0}	0.292	-0.83	0.070	-0.27	0.063	-0.85	0.056	-0.57	0.088

* evaluated at $TPR_{t_0} = 0.9$.

** MSE is in unit of 1e-2.

mNCC Under dependent censoring: Finite sample performance of the proposed one-step EM estimators, eBLUE estimators and NPMLEs as well as the IPW with full cohort size $N = 1000$ and 1 matched control. Shown below are percent in bias (PBias) and mean square error (MSE).

Table 4

(a) Accuracy estimates at $c_p = \mathcal{F}^{-1}(p)$ for $p = 0.25, 0.5, 0.75$.											
	one-step			eBLUE			NPMLE			IPW	
	True	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE
$\text{TPR}_{t_0}^{(c, .25)}$	0.953	0.15	0.021	0.51	0.021	0.32	0.014	-0.01	0.033		
$\text{TPR}_{t_0}^{(c, .5)}$	0.834	0.55	0.091	1.27	0.094	0.94	0.071	0.06	0.137		
$\text{TPR}_{t_0}^{(c, .75)}$	0.602	1.22	0.183	2.26	0.186	2.00	0.158	-0.11	0.287		
$\text{FPR}_{t_0}^{(c, .25)}$	0.715	1.04	0.290	2.03	0.283	3.18	0.199	0.53	0.366		
$\text{FPR}_{t_0}^{(c, .5)}$	0.443	1.85	0.252	2.73	0.245	5.96	0.214	0.75	0.327		
$\text{FPR}_{t_0}^{(c, .75)}$	0.190	2.77	0.119	3.08	0.116	9.89	0.115	0.35	0.152		
$\text{PPV}_{t_0}^{(c, .25)}$	0.185	1.20	0.035	1.27	0.037	0.64	0.029	0.43	0.040		
$\text{PPV}_{t_0}^{(c, .5)}$	0.243	1.06	0.070	0.94	0.069	-0.82	0.055	0.61	0.076		
$\text{PPV}_{t_0}^{(c, .75)}$	0.350	1.23	0.181	1.39	0.181	-2.08	0.144	1.28	0.192		
$\text{NPV}_{t_0}^{(c, .25)}$	0.973	-0.04	0.005	-0.02	0.005	-0.14	0.005	-0.03	0.006		
$\text{NPV}_{t_0}^{(c, .5)}$	0.952	-0.04	0.009	-0.02	0.009	-0.17	0.008	-0.03	0.009		

(a) Accuracy estimates at $c_p = \mathcal{F}^{-1}(\rho)$ for $p = 0.25, 0.5, 0.75$.

	one-step			eBLUE			NPMLE			IPW		
	True	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE	MSE
$NPV_{t_0}(c_{.75})$	0.923	-0.07	0.012	-0.05	0.014	-0.22	0.011	-0.05	0.013	-0.05	0.013	0.013

(b) Accuracy estimates at $TPR_{t_0} = 0.9, AUC_{t_0}$ and DMR_{t_0}

	one-step			eBLUE			NPMLE			IPW		
	True	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE	PBias	MSE	MSE
$FPR_{t_0}^*$	0.568	-1.16	0.218	-1.31	0.220	1.50	0.193	-1.32	0.271	-1.32	0.271	0.271
$PPV_{t_0}^*$	0.212	2.59	0.051	2.46	0.051	1.69	0.045	1.73	0.071	1.73	0.071	0.071
$NPV_{t_0}^*$	0.962	-0.18	0.004	-0.19	0.004	-0.31	0.004	-0.16	0.005	-0.16	0.005	0.005
AUC_{t_0}	0.786	-0.93	0.065	-0.64	0.059	-1.89	0.072	-0.91	0.076	-0.91	0.076	0.076
DMR_{t_0}	0.292	-0.86	0.057	-0.08	0.055	-3.59	0.065	-1.23	0.067	-1.23	0.067	0.067

* evaluated at $TPR_{t_0} = 0.9$.

** MSE is in unit of 1e-2.

Accuracy measures estimates (Est) and standard errors (SE) of DCP from HALTC nested-case-control study using one-step hybrid estimators, eBLUE, NPMLE and IPW at $t_0=3$.

Table 5

(a) Accuracy estimates at $c_p = F^{-1}(p)$ for $p = 0.25, 0.5, 0.75$.									
	one-step		eBLUE		NPMLE		IPW		
	Est	SE	Est	SE	Est	SE	Est	SE	SE
$TPR_{t_0}(c_{.25})$	0.914	0.027	0.924	0.026	0.880	0.021	0.924	0.027	0.027
$TPR_{t_0}(c_{.5})$	0.773	0.053	0.779	0.052	0.731	0.041	0.783	0.057	0.057
$TPR_{t_0}(c_{.75})$	0.487	0.087	0.487	0.087	0.431	0.069	0.490	0.089	0.089
$FPR_{t_0}(c_{.25})$	0.717	0.052	0.718	0.052	0.685	0.031	0.723	0.054	0.054
$FPR_{t_0}(c_{.5})$	0.482	0.060	0.482	0.060	0.463	0.038	0.479	0.063	0.063
$FPR_{t_0}(c_{.75})$	0.163	0.042	0.160	0.042	0.145	0.029	0.153	0.041	0.041
$PPV_{t_0}(c_{.25})$	0.014	0.005	0.013	0.005	0.029	0.007	0.027	0.007	0.007
$PPV_{t_0}(c_{.5})$	0.018	0.006	0.016	0.006	0.035	0.008	0.034	0.009	0.009
$PPV_{t_0}(c_{.75})$	0.033	0.013	0.027	0.013	0.065	0.019	0.064	0.019	0.019
$NPV_{t_0}(c_{.25})$	0.997	0.002	0.999	0.002	0.991	0.002	0.994	0.003	0.003
$NPV_{t_0}(c_{.5})$	0.995	0.002	0.997	0.002	0.988	0.003	0.991	0.003	0.003
$NPV_{t_0}(c_{.75})$	0.993	0.003	0.995	0.003	0.985	0.003	0.987	0.004	0.004

(b) Accuracy estimates at $TPR_{t_0} = 0.9, AUC_{t_0}$ and DMR_{t_0}

	one-step		eBLUE		NPMLE		IPW	
	Est	SE	Est	SE	Est	SE	Est	SE
$FPR_{t_0}^*$	0.688	0.059	0.669	0.058	0.714	0.019	0.663	0.058
$PPV_{t_0}^*$	0.015	0.005	0.014	0.005	0.028	0.006	0.028	0.007
$NPV_{t_0}^*$	0.996	0.002	0.998	0.002	0.992	0.002	0.993	0.002
AUC_{t_0}	0.713	0.049	0.726	0.047	0.684	0.025	0.722	0.048
DMR_{t_0}	0.236	0.050	0.265	0.044	0.214	0.026	0.245	0.046

* evaluated at $TPR_{t_0} = 0.9$.