Published in final edited form as:
Lifetime Data Anal. 2013 April ; 19(2): 142–169. doi:10.1007/s10985-012-9235-3.

Subgroup specific incremental value of new markers for risk prediction

Qian M. Zhou,
Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada, qmzhou@stat.sfu.ca

Yingye Zheng, and
Public Health Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA, yzheng@fhcrc.org

Tianxi Cai
Department of Biostatistics, Harvard University, Boston, MA 02115, USA

Abstract

In many clinical applications, understanding when measurement of new markers is necessary to provide added accuracy to existing prediction tools could lead to more cost effective disease management. Many statistical tools for evaluating the incremental value (IncV) of the novel markers over the routine clinical risk factors have been developed in recent years. However, most existing literature focuses primarily on global assessment. Since the IncVs of new markers often vary across subgroups, it would be of great interest to identify subgroups for which the new markers are most/least useful in improving risk prediction. In this paper we provide novel statistical procedures for *systematically* identifying potential traditional-marker based subgroups in whom it might be beneficial to apply a new model with measurements of both the novel and traditional markers. We consider various conditional time-dependent accuracy parameters for censored failure time outcome to assess the subgroup-specific IncVs. We provide non-parametric kernel-based estimation procedures to calculate the proposed parameters. Simultaneous interval estimation procedures are provided to account for sampling variation and adjust for multiple testing. Simulation studies suggest that our proposed procedures work well in finite samples. The proposed procedures are applied to the Framingham Offspring Study to examine the added value of an inflammation marker, C-reactive protein, on top of the traditional Framingham risk score for predicting 10-year risk of cardiovascular disease.

Keywords

Incremental value; Partial area under the ROC curve; Prognostic accuracy; Risk prediction; Subgroup analysis; Time dependent ROC analysis

1 Introduction

Risk models have been applied in medical practice for prediction of long-term incidence or progression of many chronic diseases such as cardiovascular disease (CVD) and cancer. With the advancement in science and technology, a wide range of biological and genomic

Correspondence to: Tianxi Cai.
tcai@hsph.harvard.edu.

markers have now become available to assist in risk prediction. However, due to the potential financial and medical costs associated with measuring these markers, their ability in improving the prediction of disease outcomes and treatment response over existing risk models needs to be rigorously accessed.

Effective statistical tools for evaluating the incremental value (IncV) of the novel markers over the routine clinical risk factors are crucial in the field of outcome prediction. Many of newly discovered markers, while promising and strongly associated with clinical outcomes, may have limited capacity in improving risk prediction over and above routine clinical variables (Tice et al. 2005; Wacholder et al. 2010). For example, on top of traditional risk variables from the Framingham risk score (FRS) (Wilson et al. 1998), the inflammation biomarker, C-reactive protein (CRP), was shown to provide modest prognostic information (Cook et al. 2006; Blumenthal et al. 2007; Ridker et al. 2007) while a genetic risk score consisting of 101 single nucleotide polymorphisms was reported as not useful (Paynter et al. 2010). In a recent paper, Wang et al. (2006) concluded that almost all new contemporary biomarkers for prevention of coronary heart disease added rather moderate *overall* predictive values to the FRS.

One possible explanation for the minimal improvement at the population average level is that the new markers may only be useful for certain subpopulations. For example, while much debate about the clinical utility of CRP remains, there is empirical evidence that CRP may substantially improve the prediction for subjects at intermediate risk (Ridker 2007). Such finding, if valid, would be extremely useful in clinical practice, since identifying the subgroups where markers can provide valuable improvement in prediction will not only lead to more informed clinical decisions but also reduced cost and effort compared to measuring novel markers on the entire population. However, to ensure the validity of such claims and more precisely pinpoint such specific subgroups, rigorous and systematic analytical tools for IncV evaluation are needed.

To quantify the *global* IncV of new markers for risk prediction, various approaches have been advocated. For example with the most popular one being focused on a comparison of summary measures of accuracy under a conventional and new models respectively (Heagerty and Zheng 2005; Uno et al. 2007; Cai and Cheng 2008). Excellent discussions on the choices of different accuracy measures can be found in Gail and Pfeiffer (2005). However, these measures quantify the overall IncV of new markers averaged over the entire study population and do not provide information on how the IncV may vary across different groups of subjects. If there are pre-defined subgroups, these measures could be estimated for each of the subgroups. However, in practice, it is often unclear how to optimally select subgroups for comparisons and ad-hoc subgroup analyses without careful planning and execution may lead to invalid results (Rothwell 2005; Pfeffer and Jarcho 2006; Wang et al. 2007). Furthermore, it is vitally important to adjust for multiple comparisons when conducting any subgroup analysis. Thus, an important question is how to *systematically* identify the potential subgroups who would benefit from the additional markers properly adjusting for multiple comparisons. There is a paucity of statistical literature on approaches for identifying such subgroups (D'Agostino 2006). Tian et al. (2009) proposed an inference procedure to estimate the IncVs in absolute prediction error of new markers in various subgroups of patients classified by the conventional markers. However, their method does not incorporate censoring. In addition, the subgroups in their paper were defined as groups of subjects whose conventional risk scores lie in different pre-assigned intervals. However, how to determine the length of intervals could be an issue. Uno et al. (2011b) proposed estimation procedures for the conditional quantiles of the improvement in the predicted risk separately for the cases and the controls. However, they did not provide procedures for determining which subgroups should be recommended to have the new markers measured.

Furthermore, no procedures were provided to account for the sampling variation or control overall type I error which is particularly important in subgroup analysis.

In this paper, we propose systematic approaches to analyzing censored event time data for identifying subgroups of patients for whom the new markers have the most or least IncV. We consider two common accuracy measures, the partial area under the ROC curve (pAUC) and the integrated discrimination improvement (IDI) index. Compared with the standard C-statistic, for many applications, the pAUC is often advocated as a better summary measure (Dwyer 1996; Dodd and Pepe 2003; Cai and Dodd 2008), since clinical interests often lie only in a specific range of the false positive rates (FPRs) or true positive rates (TPRs). For example, the region with low FPR is of more concern for disease screening (Baker and Pinsky 2001); while the region with high TPR is of more concern for the prognosis of serious disease (Jiang et al. 1996). However, the ROC curve does not capture certain aspects of the predicted absolute risk, since it is scale invariant. Many model performance measures, including the reclassification table (Cook and Ridker 2009), net reclassification improvement (NRI) and IDI (Pencina et al. 2008), proportion of case followed (PCF) and proportion needed to follow-up (PNF) (Pfeiffer and Gail 2010), have been proposed recently to overcome the limitation of the ROC curve method. Many of these measures, such as the reclassification table, NRI, PCF and PNF, rely on pre-specified clinically meaningful risk or quantile threshold values which may not be available for most diseases. For illustration purposes, we focus primarily on pAUC and IDI in this paper but note that our procedures can be easily extended to accommodate other accuracy measures.

The rest of paper is organized as follows. In Sect. 2, we present our proposed non-parametric estimation procedure for subgroup-specific IncV of new markers and along with their corresponding interval estimation procedures. In particular, resampling based simultaneous interval estimation procedures are provided as convenient and effective tools to control for multiple comparisons. We describe results from our simulation studies in Sect. 3 and the analyses of the Framingham Offspring Study using our proposed procedures in Sect. 4. Concluding remarks are given in Sect. 5. All the technical details are included in the appendices.

2 Methods

2.1 Risk modeling with and without new markers

Let X denote a set of conventional markers and let Z denote a set of new markers. Due to censoring, for the event time $T^\dagger$, one can only observe $T = \min(T^\dagger, C)$, $\Delta = I(T^\dagger \le C)$, where C is the censoring time, which is assumed to be independent of $T^\dagger$ conditional on (X, Z). See below for more discussions about censoring assumptions. Furthermore, define $Y^\dagger = I(T^\dagger \le t_0)$, where t_0 is the prediction time of clinical interest, and $Y = I(T \le t_0)$. Let $\mathscr{P}_1(X) = pr\left(Y^\dagger = 1 | X\right)$ and $\mathscr{P}_2(X, Z) = pr\left(Y^\dagger = 1 | X, Z\right)$ be the true conditional risk of developing the event by time t_0 conditional on X only and (X, Z), respectively. Suppose a data set for analysis consists of n independent realizations of (T, Δ, X, Z), $\{(T_i, \Delta_i, X_i, Z_i)\}$. Although $Y^\dagger$ and the conditional risk functions depend on t_0, we suppress t_0 from the notation throughout for the ease of presentation. From the Neyman–Pearson Lemma and similar arguments as given in McIntosh and Pepe (2002), it is not difficult to show that $\mathscr{P}_1(X)$ achieves the optimal ROC curve for predicting $Y^\dagger$ based on X only. Similarly, $\mathscr{P}_2(X, Z)$ is the optimal score for prediction $Y^\dagger$ given (X, Z).

To estimate $\mathscr{P}_1(X)$ and $\mathscr{P}_2(X, Z)$, one may consider a fully non-parametrical approach (Li and Doss 1995). However, in practice, such non-parametric estimates may perform poorly when the dimension of X or Z is not small due to the curse of dimensionality (Robins and

Ya'Acov 1997). An alternative feasible way is approximate $\mathscr{P}_1(\cdot)$ and $\mathscr{P}_2(\cdot)$ by imposing simple working models

$$pr\left(Y^{\dagger}=1|X\right)=g_1\left(\beta' V\right), \quad pr\left(Y^{\dagger}=1|X,Z\right)=g_2\left(\gamma' W\right), \quad (1)$$

where V, a $p \times 1$ vector, is a function of X, W, a $q \times 1$ vector, is a function of X and Z, β and γ are vectors of unknown regression parameters, and g_1 and g_2 are known, smooth, increasing functions. An estimator of β and γ can be obtained respectively by solving the following inverse probability weighted (IPW) estimating equations as given in Uno et al. (2007):

$$\sum_{i=1}^{n}\widehat{\omega}_i V_i\left\{Y_i - g_1\left(\beta' V_i\right)\right\}=0, \quad \sum_{i=1}^{n}\widehat{\omega}_i W_i\left\{Y_i - g_2\left(\gamma' W_i\right)\right\}=0. \quad (2)$$

where $\widehat{\omega}_i=\Delta_i I\left(T_i \le t_0\right)/\widehat{G}_{X_i,Z_i}\left(T_i\right)+I\left(T_i>t_0\right)/\widehat{G}_{X_i,Z_i}\left(t_0\right)$, and $\widehat{G}_{X,Z}(t)$ is a root-n consistent estimator of $G_{X,Z}(t)=pr(C\ge t|X,Z)$. This IPW estimator may be justified heuristically the argument that $E\left\{\widehat{\omega}_i I\left(Y_i=y\right)|T_i^{\dagger},X_i,Z_i\right\}\approx E\left\{I\left(Y_i^{\dagger}=y\right)|T_i^{\dagger},X_i,Z_i\right\}$, for $y=0$, 1. Let $\widehat{\beta}$ and $\widehat{\gamma}$ be the resulting estimator of β and γ, respectively. For a subject with $X=x$, $Z=z$ whose $V=v$ and $W=w$, the risk is estimated by $\widehat{p}_1(x)=g_1\left(\widehat{\beta}' v\right)$ based on X alone and by $\widehat{p}_2(x,z)=g_2\left(\widehat{\gamma}' \omega\right)$ based on X and Z. It has been previously shown in Uno et al. (2007) that regardless of the adequacy of the working model (1), $\widehat{\theta}=\left(\widehat{\beta}',\widehat{\gamma}'\right)'$ converges in probability to a deterministic vector $\theta_0=\left(\beta_0',\gamma_0'\right)'$ as $n\to\infty$. Let $\bar{p}_1(x)=g_1\left(\beta_0' v\right)$ and $\bar{p}_2(x,z)=g_2\left(\gamma_0' \omega\right)$. When the models in (1) are correctly specified, $\bar{p}_1(x)=\mathscr{P}_1(x)$ and $\bar{p}_2(x,z)=\mathscr{P}_2(x,z)$. To obtain a valid estimator $\widehat{G}_{X,Z}(\cdot)$, one may impose a semi-parametric model, such as the proportional hazards (PHs) model (Cox 1972), for $G_{X,Z}(t)$ and obtain $\widehat{G}_{X,Z}(t)$ as $\exp\left\{\widehat{\Lambda}_0(t)\exp\left(\widehat{\gamma}_c' W_c\right)\right\}$, where W_c is a function of (X, Z), $\widehat{\gamma}_c$ is the maximum partial likelihood estimator and $\widehat{\Lambda}_0(t)$ is the Breslow's estimator. When the censoring is independent of both T and (X, Z), one may obtain $\widehat{G}_{X,Z}(\cdot)$ simply as the Kaplan–Meier estimator. It is important to note that if the in (1) only hold for a given t_0 and the dimension of (X, Z) is not small, root-n consistent estimators of β and γ may not exist without imposing additional modeling assumptions about $G_{X,Z}(t)$ due to the curse of dimensionality (Robins and Ya'Acov 1997).

2.2 Subgroup specific IncVs

For illustration purposes, we consider two accuracy measures, the pAUC and the IDI index. We first define both concepts in the context of evaluating a risk score/model. Suppose that we use $\bar{p}_2(X,Z)$ as a risk score for classifying the event status $Y^{\dagger}$, and without loss of generality, we assume that a higher value of $\bar{p}_2(X,Z)$ is associated with a higher risk and refer to the two states, $Y^{\dagger}=1$ and $Y^{\dagger}=0$, as "diseased" and "disease-free" or "cases" and "controls". The discrimination capacity of $\bar{p}_2(X,Z)$ can be quantified based on the ROC curve, which is a plot of the TPR function, $\mathscr{S}_1(c)\equiv pr\left\{\bar{p}_2(X,Z)\ge c|Y^{\dagger}=1\right\}$, against the FPR function, $\mathscr{S}_0(c)\equiv pr\left\{\bar{p}_2(X,Z)\ge c|Y^{\dagger}=0\right\}$. The ROC curve, $\mathrm{ROC}(u)=\mathscr{S}_1\left\{\mathscr{S}_0^{-1}(u)\right\}$ describes the inherent capacity of distinguishing "cases" from "controls". The pAUC with a

restricted region of FPR, say FPR ≤ f, is given by $\text{pAUC}_f = \int_0^f \text{ROC}(u)\,du$, for $f \in [0, 1]$. The IDI index, is simply $\text{IDI} = \int_0^1 \mathscr{S}_1(c)\,dc - \int_0^1 \mathscr{S}_0(c)\,dc$.

To evaluate how the IncV of Z may vary across subgroups defined by X, we define new conditional pAUC and IDI index. We propose to use $\bar{p}_1(x)$ as a scoring system for grouping subjects with potentially similar initial risk estimates and create subgroups $\upsilon_s = \{X : \bar{p}_1(X) = s\}$. Then we evaluate the IncV of Z for each υ_s based on how well $\bar{p}_2(X, Z)$ can further discriminate subjects within υ_s with $Y^\dagger = 1$ from those with $Y^\dagger = 0$. Suppose $\bar{p}_2(X, Z)$ is used to classify $Y^\dagger$ for subjects in υ_s. The TPR and FPR of the classification rule $\bar{p}_2(X, Z) \geq c$ given υ_s $\mathscr{S}_1(c;s)$ and $\mathscr{S}_0(c;s)$, respectively, where

$$\mathscr{S}_y(c;s) = pr\left\{\bar{p}_2(X,Z) \geq c | Y^\dagger = y, \quad \bar{p}_1(X) = s\right\}, \quad \text{and} \quad c \in (0,1), \quad y = 0, 1.$$

Conditional on $\bar{p}_1(X) = s$, the ROC curve of $\bar{p}_2(X, Z)$ is $\text{ROC}(u;s) = \mathscr{S}_1\left\{\mathscr{S}_0^{-1}(u;s);s\right\}$, for $u \in [0, 1]$. The conditional pAUC is given by $\text{pAUC}_f(s) = \int_0^f \text{ROC}(u;s)\,du$, $f \in [0,1]$. Note that $f = 1$ yields conditional AUC(s. If Z is non-informative for υ_s, the corresponding ROC curve would be a diagonal line, and we expect that $\text{pAUC}_s = f^2/2$, which is the area under a diagonal line. Thus, the subgroup υ_s specific IncV of Z with respect to (wrt) the pAUC is given by $\text{pAUC}_f(s) - f^2/2$. The IDI index conditional on $\bar{p}_1(X) = s$ is given by

$$\begin{aligned} \text{IDI}(s) &= \int_0^1 \mathscr{S}_1(c;s)\,dc - \int_0^1 \mathscr{S}_0(c;s)\,dc \\ &= E\left\{\bar{p}_2(X,Z) | Y^\dagger = 1, \quad \bar{p}_1(X) = s\right\} - E\left\{\bar{p}_2(X,Z) | Y^\dagger = 0, \quad \bar{p}_1(X) = s\right\}. \end{aligned} \tag{3}$$

If Z is non-informative for this subgroup υ_s, the conditional IDI index would be 0, and therefore, the subgroup υ_s specific IncV of Z wrt the IDI index is IDI(s). Based on these subgroup-specific IncVs, we are able to identify the set of s such that Z is useful to improve the prediction accuracy for υ_s, which is referred to as the *effective subpopulation* υ^* in our paper. Specifically, the effective subpopulation wrt pAUC is defined as $\upsilon^* = \{X : \text{pAUC}_f(\bar{p}_1(X)) - f^2/2 > c_1\}$; the effective subpopulation wrt IDI are defined as $\upsilon^* = \{X : \text{IDI}(\bar{p}_1(X)) > c_2\}$, where c_1 and c_2 are some possibly data dependent threshold values. For the subjects in the effective subpopulation, measurement of new markers would provide added accuracy to the conventional risk model.

2.3 Inference about subgroup-specific IncVs

We first discuss the estimation for the conditional TPR and FPR functions $\left\{\mathscr{S}_y(c;s), y = 0, 1\right\}$ since both pAUC$_f$(s) and IDI(s) are simple functionals of these two functions. Let $\widehat{p}_{1i} = \widehat{p}_1(X_i) = g_1\left(\widehat{\beta}' V_i\right)$ and $\widehat{p}_{2i} = \widehat{p}_2(X_i, Z_i) = g_2\left(\widehat{\gamma}' W_i\right)$. To obtain a consistent estimator of $\mathscr{S}_y(c;s)$, since $\mathscr{S}_y(C;s)$ is between 0 and 1, we consider a non-parametric local likelihood estimation method (Tibshirani and Hastie 1987) along with IPW accounting for censoring. Specifically, we obtain $\left\{\widehat{a}_{y,h_y}(c;s), \widehat{b}_{y,h_y}(c;s)\right\}$ as the solution to the IPW local likelihood score equation,

$$\sum_{i:Y_i=y} \widehat{\omega}_i \begin{bmatrix} 1 \\ h_y^{-1}\widehat{\varepsilon}_{1i}(s) \end{bmatrix} K_{h_y}\{\widehat{\varepsilon}_{1i}(s)\}\left[I(\widehat{p}_{2i} \geq c) - g\{a + b\widehat{\varepsilon}_{1i}(s)\}\right] = 0, \tag{4}$$

where $\widehat{\varepsilon}_{1i}(s) = \phi(\widehat{p}_{1i}) - \phi(s)$, $g(x) = \exp(x)/\{1 + \exp(x)\}$, $K_h(x) = K(x/h)/h$, $K(\cdot)$ is a known smooth symmetric kernel density function with a bounded support, and the bandwidth $h_y > 0$ is assumed to be $O(n^{\nu})$, for $1/5 < \nu < 1/2$, and $\varphi(\cdot)$ is a known, non-decreasing transformation function that can potentially be helpful in improving the performance of the smoothed estimator (Wand et al. 1991; Park et al. 1997). Then, $\mathscr{S}(c;s)$ can be estimated by $\widehat{\mathscr{S}}_{y,h_y}(c;s) = g\left\{\widehat{a}_{y,h_y}(c;s)\right\}$ for $y = 0, 1$. In the Appendix A.1, we show that $\widehat{S}_{y,h_y}(c;s) - \mathscr{S}_y(c;s) \to 0$ in probability as $n \to \infty$, uniformly in $c \in [0, 1]$ and $s \in \mathscr{I}_{h_y} \equiv \left[\phi^{-1}\left(\rho_l = h_y\right), \phi^{-1}\left(\rho_u - h_y\right)\right]$ where $[\rho_l, \rho_u]$ is a subset of the support of $\phi\left\{g_1\left(\beta_0' V\right)\right\}$ and β_0 is the limit of $\widehat{\beta}$. As a special case, by setting b in (4) to 0, one may obtain a local constant estimator,

$$\widehat{\mathscr{S}}_{y,h_y}(c;s) = \frac{\sum_{i=1}^{n} \widehat{\omega}_i I(Y_i = y) K_{h_y}\{\widehat{\varepsilon}_{1i}(s)\} I(\widehat{p}_{2i} \geq c)}{\sum_{i=1}^{n} \widehat{\omega}_i I(Y_i = y) K_{h_y}\{\widehat{\varepsilon}_{1i}(s)\}}, \quad y = 0, 1.$$

2.4 Inference procedures for $\mathrm{pAUC}_f(s)$

Based on $\widehat{\mathscr{S}}_{y,h_y}(c;s)$ can be estimated as

$$\widehat{\mathrm{pAUC}}_f(s) = \int_0^f \widehat{\mathrm{ROC}}(u;s)\, du = \int_{\widehat{\mathscr{S}}_{0,h_0}^{-1}(f;s)}^{\infty} \widehat{\mathscr{S}}_{1,h_1}(c;s)\, d\left\{1 - \widehat{\mathscr{S}}_{0,h_0}(c;s)\right\}.$$

where $\widehat{\mathrm{ROC}}(u;s) = \widehat{\mathscr{S}}_{1,h_1}\left\{\widehat{\mathscr{S}}_{0,h_0}^{-1}(u;s);s\right\}$ and (h_0, h_1) is the pair of optimal band-widths for estimating $\mathscr{S}_0(c;s)$ and $\mathscr{S}_1(c;s)$, respectively. In the Appendix A.2, we show that $\widehat{\mathrm{pAUC}}_f(s)$ is uniformly consistent for $\mathrm{pAUC}_f(s)$.

As a special case, when both X and Z are univariate, the ROC curve of $\bar{p}_2(X, Z)$ conditional on $\bar{p}_1(X)$ is equivalent to the ROC curve of Z conditional on X since the ROC curve is scale invariant. A simple local constant IPW estimator of $\mathscr{S}_y(z;x)$ is given by

$$\widehat{\mathscr{S}}_{y,h_y}(z;x) = \frac{\sum_{i=1}^{n} \widehat{\omega}_i I(Y_i = y) K_{h_y}(X_i - x) I(Z_i \geq z)}{\sum_{i=1}^{n} \widehat{\omega}_i I(Y_i = y) K_{h_y}(X_i - x)}.$$

The resulting estimator of $\mathrm{pAUC}_f(x)$ is

$$\begin{aligned} \widehat{\mathrm{pAUC}}_f(x) &= \int_0^f \widehat{\mathscr{S}}_{1,h_1}\left\{\widehat{\mathscr{S}}_{0,h_0}^{-1}(u;x);x\right\} du \\ &= \frac{\sum_{i=1}^{n} \widehat{\omega}_i Y_i K_{h_1}(X_i - x) U_i(x;f,h_0)}{\sum_{i=1}^{n} \widehat{\omega}_i Y_i K_{h_1}(X_i - x)}, \end{aligned}$$

where $U_i(x;f,h_0) = f - \min\left\{f, \widehat{\mathscr{S}}_{0,h_0}(Z_i;x)\right\}$ is the estimated truncated placement value proposed by Cai and Dodd (2008).

It is difficult to directly estimate the variance of $\widehat{\mathscr{W}}_{\mathrm{pAUC}_f}(s) = \sqrt{nh_1}\left\{\widehat{\mathrm{pAUC}}_f(s) - \mathrm{pAUC}_f(s)\right\}$ since it involves unknown derivative functions. We propose a perturbation-resampling method to approximate the distribution of $\widehat{\mathscr{W}}_{\mathrm{pAUC}f}(s)$. This method has been widely used in survival analyses (see for example, Jin et al. 2001; Park and Wei 2003; Cai et al. 2005). To be specific, let $\Xi = \{\xi_i, i = 1, \ldots, n\}$ be n independent positive random variables following a known distribution with mean 1 and variance 1, and Ξ is independent of the data. For each set of Ξ, we first obtain β^* and γ^*, as the respective solutions to

$$\sum_{i=1}^{n}\omega_i^* V_i\left\{Y_i - g_1\left(\beta' V_i\right)\right\}\xi_i = 0, \quad \text{and} \quad \sum_{i=1}^{n}\omega_i^* W_i\left\{Y_i - g_2\left(\gamma' W_i\right)\right\}\xi_i = 0,$$

where $\omega_i^* = \Delta_i I(T_i \le t_0)/G^*_{x_i,Z_i}(T_i) + I(T_i > t_0)/G^*_{X_iZ_i}(t_0)$ and $G^*_{X,Z}$ is the perturbed estimator of $G_{X,Z}(\cdot)$ with Ξ being the weights. Let $p^*_{1i} = g_1\left(\beta^{*\prime} V_i\right)$, $\varepsilon^*_{1i}(s) = \phi\left(p^*_{1i}\right) - \phi(s)$, $p^*_{2i} = g_2\left(\gamma^{*\prime} W_i\right)$, and $M_i^*(c) = I\left(p^*_{2i} \ge c\right)$. Subsequently, we obtain the perturbed counterpart of $\mathscr{S}_y(c;s)$ as $\mathscr{S}^*_{y,h_y}(c;s) = g\left\{a^*_{y,h_y}(c;s)\right\}$, where $a^*_{y,h_y}(c;s)$ is the solution to the perturbed score equation

$$\sum_{i=1}^{n}\omega_i^* I(Y_i = y)\begin{bmatrix} 1 \\ h_y^{-1}\varepsilon^*_{1i}(s) \end{bmatrix} K_{h_y}\left\{\varepsilon^*_{1i}(s)\right\}\left[M_i^*(c) - g\left\{a + b\varepsilon^*_{1i}(s)\right\}\right]\xi_i = 0.$$

Then, the perturbed pAUC is given by, $\mathrm{pAUC}^*_f(s) = \int_0^f \mathrm{ROC}^*(u;s)\,du$, where $\mathrm{ROC}^*(u;s) = \mathscr{S}^*_{1,h_1}\left\{\mathscr{S}^{*-1}_{0,h_0}(u;s);s\right\}$. In the Appendix A.3, we show that the unconditional distribution of $\widehat{\mathscr{W}}_{\mathrm{pAUC}f}(s)$ can be approximated by the conditional distribution of

$$W^*_{\mathrm{pAUC}_f}(s) = \sqrt{nh_1}\left\{\mathrm{pAUC}^*_f(s) - \widehat{\mathrm{pAUC}}_f(s)\right\}, \quad (5)$$

given the data. With the above resampling method, for any fixed $s \in \mathscr{I}_{h_1}$, one may obtain a variance estimator of $\widehat{\mathscr{W}}_{\mathrm{pAUC}_f}(s)$, $\widehat{\sigma}^2_f(s)$, based on the empirical the variance of B realizations from (5). For any fixed $s \in \mathscr{I}_{h_1}$ and $\alpha \in (0, 1)$, a pointwise $100(1-\alpha)\%$ confidence interval (CI) for $\mathrm{pAUC}_f(s)$ can be constructed via $\widehat{\mathrm{pAUC}}_f(s) \pm (nh_1)^{-1/2} c_\alpha \widehat{\sigma}_f(s)$, where c_α is the $100(1-\alpha)$th percentile of the standard normal distribution.

2.4.1 Inference for IDI(s)—Based on $\widehat{\mathscr{S}}_{y,h_y}(c;s)$, we may obtain plug-in estimators for $\mathrm{IS}(s) = \int_0^1 \mathscr{S}_1(c;s)\,dc$ and $\mathrm{IP}(s) = \int_0^1 \mathscr{S}_0(c;s)\,dc$ respectively as

$$\widehat{\mathrm{IS}}(s) = \int_0^1 \widehat{\mathscr{S}}_{1,h_1}(c;s)\,dc, \quad \text{and} \quad \widehat{\mathrm{IP}}(s) = \int_0^1 \widehat{\mathscr{S}}_{0,h_0}(c;s)\,dc.$$

Thus, IDI(s) can be estimated by $\widehat{\mathrm{IDI}}(s) = \widehat{\mathrm{IS}}(s) - \widehat{\mathrm{IP}}(s)$. Similar to the derivations given in the Appendix A for $\widehat{\mathrm{pAUC}}_f(s)$, the asymptotic results for $\widehat{S}_{y,h_y}(c;s)$ can be directly used to establish the consistency and asymptotic normality for $\widehat{\mathrm{IDI}}(s)$. In addition, the unconditional distribution of $\widehat{\mathscr{W}}_{\mathrm{IDI}}(s) = \sqrt{nh_1}\left\{\widehat{\mathrm{IDI}}(s) - \mathrm{IDI}(s)\right\}$ can be approximated the conditional

distribution of $\mathscr{W}^*_{\mathrm{IDI}}(s) = \sqrt{nh_1}\left\{\mathrm{IDI}^*(s) - \widehat{\mathrm{IDI}}(s)\right\}$, given the data, where $\mathrm{IDI}^*(s) = \int_0^1 \mathscr{S}^*_{1,h_1}(c;s)\,dc - \int_0^1 \mathscr{S}^*_{0,h_0}(c;s)\,dc$ and $\mathscr{S}^*_{y,h_y}(c;s)$ is the perturbed counterpart of $\mathscr{S}_{y,h_y}(c;s)$, y=0, 1. The pointwise CIs for any fixed $s \in \mathscr{I}_{h_1}$ are constructed in a similar way as the inference for pAUC$_I(s)$. As a special case, a kernel local constant estimator of IDI(s) is given by

$$\widehat{\mathrm{IDI}}(s) = \frac{\sum_{i=1}^n \widehat{\omega}_i Y_i K_{h_1}\{\widehat{\varepsilon}_{1i}(s)\}\widehat{p}_{2i}}{\sum_{i=1}^n \widehat{\omega}_i Y_i K_{h_1}\{\widehat{\varepsilon}_{1i}(s)\}} - \frac{\sum_{i=1}^n \widehat{\omega}_i (1-Y_i) K_{h_0}\{\widehat{\varepsilon}_{1i}(s)\}\widehat{p}_{2i}}{\sum_{i=1}^n \widehat{\omega}_i (1-Y_i) K_{h_0}\{\widehat{\varepsilon}_{1i}(s)\}},$$

with the perturbed counterpart given by

$$\mathrm{IDI}^*(s) = \frac{\sum_{i=1}^n \omega_i^* Y_i K_{h_1}\left\{\varepsilon^*_{1i}(s)\right\} p^*_{2i}\xi_i}{\sum_{i=1}^n \omega_i^* Y_i K_{h_1}\left\{\varepsilon^*_{1i}(s)\right\}\xi_i} - \frac{\sum_{i=1}^n \omega_i^* (1-Y_i) K_{h_0}\left\{\varepsilon^*_{1i}(s)\right\} p^*_{2i}\xi_i}{\sum_{i=1}^n \omega_i^* (1-Y_i) K_{h_0}\left\{\varepsilon^*_{1i}(s)\right\}\xi_i}.$$

Selection of the optimal bandwidths for pAUC$_I(s)$ and IDI(s) is illustrated in the Appendix B.

2.4.2 Identifying the effective subpopulation—To identify the effective subpopulation, one may simultaneously assess the subgroup-specific IncV wrt a certain accuracy measure, denoted by $\mathbb{A}(s)$, for example pAUC$_I(s) - I^2/2$ or IDI(s), over a range of s values by constructing simultaneous CI for $\{\mathbb{A}(s), s \in \mathscr{I}_{h_1}\}$. Unfortunately, the distribution of $\widehat{\mathscr{W}_{\mathbb{A}}}(s)$ does not converge as a process in s, as $n \to \infty$. Thus, we cannot apply the standard large sample theory for stochastic processes to approximate the distribution of $\widehat{\mathscr{W}_{\mathbb{A}}}(s)$. Nevertheless, by the strong approximation argument and extreme value limit theorem (Bickel and Rosenblatt 1973), we show in the Appendix A.3 that a standardized version of the sup-statistic $\Gamma = \sup_{s \in \mathscr{I}_{h_1}} |\widehat{\mathscr{W}_{\mathbb{A}}}(s)/\widehat{\sigma}_{\mathbb{A}}(s)|$ converges in distribution to a proper random variable, where $\widehat{\sigma}^2_{\mathbb{A}}$ denotes the variance estimator of $\widehat{\mathscr{W}_{\mathbb{A}}}(s)$. In practice, for large n, one can approximate the distribution of Γ based on realizations of $\Gamma^* = \sup_{s \in \mathscr{I}_{h_1}} |\mathscr{W}^*_{\mathbb{A}}(s)/\widehat{\sigma}_{\mathbb{A}}(s)|$, where $\widehat{\mathscr{W}^*_{\mathbb{A}}}$ is the perturbed counterpart of $\widehat{\mathscr{W}_{\mathbb{A}}}$. Therefore, a 100(1 − α) % simultaneous CI for $\mathbb{A}(s)$ can be obtained as $\widehat{\mathbb{A}}(s) \pm (nh_1)^{-1/2} d_\alpha \widehat{\sigma}_{\mathbb{A}}(s)$, where d_α is the empirical 100(1 − α)th quantile of Γ^*. Thus, to account for sampling variation and multiple testing, the effective subpopulation is chosen as $\left\{X : \widehat{\mathbb{A}}(\widehat{p}_1(X)) > (nh_1)^{-1/2} d_\alpha \widehat{\sigma}_{\mathbb{A}}(\widehat{p}_1(X))\right\}$ in real data analyses.

2.4.3 Test for heterogeneous IncV—Another question of interest is whether the subgroup-specific V $\mathbb{A}(s)$, for example pAUC$_I(s)$, is constant across different values of s over a certain interval $[s_l, s_u]$. We define the average IncV over $[s_l, s_u]$ as

$$\mathbb{A}_{[s_l, s_u]} = \frac{\int_{s_l}^{s_u} \mathbb{A}(s)\, d\mathscr{F}(s)}{[\mathscr{F}(s_u) - \mathscr{F}(s_l)]}$$

where $\mathscr{F}(s)=\Pr\left\{\bar{p}_1(X)\le s\right\}$, and we define the relative subgroup-specific IncV over $[s_l, s_u]$ as $\mathbb{D}_{\mathbb{A}_{[s_l,s_u]}}(s)=\mathbb{A}(s)-\mathbb{A}_{[s_l,s_u]}$. The point estimate of $\mathbb{D}_{\mathbb{A}_{[s_l,s_u]}}(s)$ is given by

$$\widehat{\mathbb{D}}_{\mathbb{A}_{[s_l,s_u]}}(s)=\widehat{\mathbb{A}}(s)-\widehat{\mathbb{A}}_{[s_l,s_u]},\quad \widehat{\mathbb{A}}_{[s_l,s_u]}=\frac{n^{-1}\sum_{i=1}^{n}\widehat{\mathbb{A}}(\widehat{p}_{1i})\,I(\widehat{p}_{1i}\in[s_l,s_u])}{\widehat{\mathscr{F}}(s_u)-\widehat{\mathscr{F}}(s_l)}$$

where $\widehat{\mathscr{F}}(s)=n^{-1}\Sigma_{i=1}^{n}I\{\widehat{p}_{1i}\le s\}$. In addition, the unconditional distribution of $\widehat{\mathscr{W}}_{\mathbb{D}}(s)=\sqrt{nh_1}\left\{\mathbb{D}_{\mathbb{A}_{s_l,s_u}}(s)-\mathbb{D}_{\mathbb{A}_{[s_l,s_u]}}(s)\right\}$ can be approximated by the conditional distribution of $\mathscr{W}^{*}_{\mathbb{D}}(s)=\sqrt{nh_1}\left\{\mathbb{D}^{*}_{\mathbb{A}_{[s_l,s_u]}}(s)-\widehat{\mathbb{D}}_{\mathbb{A}_{[s_l,s_u]}}(s)\right\}$ given the data, where $\mathbb{D}^{*}_{\mathbb{A}_{[s_l,s_u]}}(s)=\mathbb{A}^{*}(s)-\mathbb{A}^{*}_{[s_l,s_u]}$ with $\mathbb{A}^{*}(s)$ as the perturbed counterpart of $\widehat{\mathbb{A}}(s)$ and $\mathbb{A}^{*}_{[s_l,s_u]}=n^{-1}\Sigma_{i=1}^{n}\mathbb{A}^{*}(\widehat{p}_{1i})\,I(\widehat{p}_{1i}\in[s_l,s_u])/\left[\widehat{\mathscr{F}}(s_u)-\widehat{\mathscr{F}}(s_l)\right]$. The variance estimator $\widehat{\sigma}^2_{\mathbb{D}}$ of $\widehat{\mathscr{W}}_{\mathbb{D}}(s)$ can be obtained from realizations of $\mathscr{W}^{*}_{\mathbb{D}}(s)$.

If the subgroup-specific IncV of Z is constant over $[s_l, s_u]$, i.e., $\mathbb{A}(s)\equiv c_0$ for $s\in[s_l, s_u]$, $\mathbb{A}_{[s_l,s_u]}=c_0$ and $\mathbb{D}_{\mathbb{A}_{[s_l,s_u]}}(s)=0$ for $s\in[s_l, s_u]$. Testing whether the subgroup specific IncV is constant over $[s_l, s_u]$ is the equivalent to testing the null hypothesis $H_0:\mathbb{D}_{\mathbb{A}_{[s_l,s_u]}}(s)=0$ for $s\in[s_l, s_u]$. To adjust for multiple testing, we consider the standard version of the sup-statistic $\Gamma_{\mathbb{D}}=\sup_{s\in[s_l,s_u]}|\widehat{\mathscr{W}}^{(0)}_{\mathbb{D}}(s)/\widehat{\sigma}_{\mathbb{D}}(s)|$ where $\widehat{\mathscr{W}}^{(0)}_{\mathbb{D}}(s)=\sqrt{nh_1}\widehat{\mathbb{D}}_{\mathbb{A}_{[s_l,s_u]}}(s)$ is the statistic $\widehat{\mathscr{W}}_{\mathbb{D}}(s)$ under the null hypothesis H_0. One may approximate the distribution of $\Gamma_{\mathbb{D}}$ based on realizations of $\Gamma^{*}_{\mathbb{D}}=\sup_{s\in[s_l,s_u]}|\mathscr{W}^{*}_{\mathbb{D}}(s)/\widehat{\sigma}_{\mathbb{D}}(s)|$. The empirical p value for testing the null hypothesis H_0 can be obtained by $B^{-1}\Sigma_{b=1}^{B}I\left\{\Gamma^{*(b)}_{\mathbb{D}}>\Gamma_{\mathbb{D}}\right\}$, where $\left\{\Gamma^{*(b)}_{\mathbb{D}}, b=1,\ldots,B\right\}$ are B realizations of $\Gamma^{*}_{\mathbb{D}}$.

3 Simulation studies

To examine the finite sample properties of the proposed estimation procedure, we conduct a simulation study where the conventional marker X and the new marker Z are both univariate and jointly generated from a bivariate normal distribution

$$\begin{pmatrix} X \\ Z \end{pmatrix}\sim BVN\left(\begin{bmatrix} \mu_X \\ \mu_Z \end{bmatrix},\begin{bmatrix} \sigma_X^2 & \sigma_X\sigma_{Z\rho XZ} \\ \sigma_X\sigma_{Z\rho XZ} & \sigma_Z^2 \end{bmatrix}\right).$$

In this simulation study, $\mu_X=\mu_Z=0$, $\sigma_X=2$ and σ_Z 0.5, and ρ_{XZ} 0.01. The failure time T, given the markers X and Z, is generated from an accelerated failure time model with a log-normal distribution for T, i.e., $\log T=h(X, Z)+\in$, where $\in$ is a normal random variable with mean 0 and standard deviation $\sigma_T=1.5$. In this simulation study, $h(X, Z)$ is a linear model, i.e., $h(X, Z)=-\beta_X X-\beta_Z Z-\beta_{XZ}XZ$. We consider a practical situation where the new marker Z may make a major contribution to the underlying mechanism in contrast with the conventional marker X, although it may not be measured routinely. Thus, in this simulation study, we set $\beta_X=0.01$ and $\beta_Z=\beta_{XZ}=1$. The censoring time C is generated from an exponential distribution with rate c_0^{-1}. A value of $c_0\approx 20$ is chosen such that roughly

80% of the failure time is censored. A time point $t_0 \approx 0.2$ is set such that the proportion of the "cases" in the sample is approximately 20%.

We investigate the kernel local constant estimator for the conditional pAUC_f with $f = 0.1$ representing a low FPR region and $f = 1$ representing the standard AUC. Since Z and $\log T$ are jointly normal conditional on $X = x$, it is straightforward to calculate the true values of $\text{pAUC}_f(x)$. We consider a relatively smaller sample size 1,000, a moderate sample size of 5,000 and a relatively larger sample size of 10,000. Both of the pAUC with FPR ≤ 0.1, i.e., $\text{pAUC}_{0.1}(x)$ and the full AUC are estimated at a sequence of values of X. For ease of computation, the pair of the bandwidths (h_0, h_1) for constructing the non-parametric estimate was fixed at (i) for the full AUC, (2.531, 2.102) for $n = 1{,}000$, (1.905,1.534) for $n = 5{,}000$, and (1.640, 1.380) for $n = 10{,}000$; (ii) for the $\text{pAUC}_{0.1}$, (2.377,2.432) for $n = 1{,}000$, (1.843,2.361) for $n = 5{,}000$, and (1.507, 2.085) for $n = 10{,}000$. Here, $\left(h_0^{\text{opt}}, h_1^{\text{opt}}\right)$ were chosen as the average of the bandwidths selected from 10 independent simulated datasets using the two-stage of five-fold cross-validation method described in the Appendix B and $n^{-0.1}$ was multiplied to h_y^{opt} to yield the final bandwidths used for simulation. In addition, the kernel function $K(\cdot)$ was chosen as the Epanechnikov kernel. Here, since we assume that the censoring time C is independent of both T and (X, Z), $G_{X,Z}(t) = G(t)$ is estimated by a Kaplan–Meier estimator.

The performance of the point estimates and pointwise 95% CIs obtained by the resampling method was assessed from 1,000 independent replicates. For all of these scenarios, the non-parametric estimators have substantially small biases, the estimated standard errors are close to their empirical counterparts, and empirical coverage levels are close to the nominal level. In Fig. 1, we summarize the performance of the point and interval estimates for $\text{pAUC}_{0.1}$ with sample size 10,000. For this scenario, the empirical coverage probabilities of the 95% pointwise CIs range from 92.9 to 95.4%. The empirical coverage levels of the 95% simultaneous confidence bands for the standard AUC are 93.2% for $n = 1{,}000$, 93.3% for $n = 5{,}000$, and 94.5% for $n = 10{,}000$; the empirical coverage levels of the 95% simultaneous confidence bands for the $\text{pAUC}_{0.1}$ are 93.3% for $n = 1{,}000$, 93.4% for $n = 5{,}000$, and 92.5% for $n = 10{,}000$.

4 Example: the Framingham Offspring Study

The Framingham Offspring Study was established in 1971 with 5,124 participants who were monitored prospectively on epidemiological and genetic risk factors of CVD. Here, we use data from 1,687 female participants of which 261 have either died or experienced a CVD event by the end of follow-up period, and the 10-year event rate is 6%. The Framingham risk model, based on several clinical risk factors including age, systolic blood pressure, diastolic blood pressure, total cholesterol, high-density lipoprotein cholesterol, current smoking status and diabetes, is widely used in clinical settings but only with moderate accuracy for predicting the 10-year risk of CVD (Cook et al. 2006). The FRS is constructed as the weighted average of the risk factors in the Framingham risk model using β-coefficients given in Table 6 of Wilson et al. (1998). The risk estimates are obtained from the FRS through the transformation $1 - \exp\{-\exp(\cdot)\}$. The density plot of the risk estimates obtained from the FRS is shown in Fig. 2a. The overall gain in C-statistic by adding the CRP on top of FRS is 0.002 (from 0.776 to 0.778, with 95% CI (−0.005,0.01)). Note that a log transformation is applied on the CRP throughout the analysis. According to the Framingham risk model (Wilson et al. 1998) and the risk threshold values employed by the Adult Treatment Panel III of the National Cholesterol Eduction Program (Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults 2001), these 1,687 female participants may be classified into three risk groups: 1,462 as low risk (<10

%); 193 as intermediate risk (between 10% and 20%); 32 as high risk (>20 %). The IncVs wrt C-statistic are 0.00057 (with 95% CI (−0.012,0.013)) for the low risk group; 0.037 (with 95% CI (−0.054,0.13)) for the intermediate risk group; 0.034 (with 95% CI (−0.097,0.16)) for the high risk group. Note that the low risk group consists of about 87% of the entire cohort. Now we further classify the 1,462 patients of the low risk group into 10 finer subgroups with the length of the risk interval for each subgroup being 0.01, for example, 0–0.01, 0.01–0.02, and etc. The IncVs wrt C-statistics for these 10 subgroups of low risk as well as the intermediate and high risk groups with their 95% CIs are shown in Fig. 2b. This suggests that adding CRP on top of FRS may be most useful for the risk groups around 5%, which is also referred to as the intermedium low risk group in some literature.

First, we investigate the IncV of the CRP over the FRS wrt AUC, $pAUC_{0.1}$ and IDI in predicting the 10-year risk of CVD events among subgroups defined by the FRS. For the purpose of kernel smoothing, the transformation function $\phi(\cdot)$ in the local likelihood score equation (4) is $\phi(x) = \Phi\left(\frac{x-\mu X}{\sigma X}\right)$, where $\mu_X = -3.74$ is the sample mean of the FRS and $\sigma_X = 1.35$ is the sample standard deviation of the FRS, and $\Phi(x)$ is the cumulative distribution function of a standard normal distribution. Here we use local kernel constant estimates with Epanechnikov kernel. The optimal bandwidths $\left(h_0^{\text{opt}}, h_1^{\text{opt}}\right)$ in ϕ-scale are chosen via a 10-fold cross validation procedure: (0.117,0.393) for the standard AUC, (0.264,0.721) for $pAUC_{0.1}$, and (0.018,0.273) for IDI. The point estimates along with the 95% pointwise and simultaneous CIs for the subgroup-specific IncV wrt AUC, $pAUC_{0.1}$ and IDI are shown in Fig. 3. The point estimate for IDI is obtained via a cross-validation procedure to correct for biases due to overfitting. Based on the pointwise CIs of the subgroup-specific IncV wrt AUC, the addition of CRP appears to improve risk prediction for subjects with the FRS risk ranging from 0.028 to 0.096. The corresponding range is 0.008–0.148 when based on the CIs for the subgroup-specific IncV wrt $pAUC_{0.1}$; 0.004–0.102 when based on the CIs for the subgroup-specific IncV wrt IDI. After controlling for the overall type I error, inclusion of CRP may significantly improve discrimination for subjects with the FRS risk ranging from 0.034 to 0.070 based on AUC; from 0.010 to 0.078 based on $pAUC_{0.1}$; from 0.032 to 0.068 based on IDI. The IDI findings and the pAUC findings agree with each other. These results suggest that CPR might be useful to improve risk prediction among patients regarded as having low to moderate risk according to the FRS.

It is worth to note that the bandwidth selection procedure is not sensitive towards the choice of the number of folds in cross-validation. Using a five-fold cross-validation, the optimal bandwidths $\left(h_0^{\text{opt}}, h_1^{\text{opt}}\right)$ are (0.121, 0.394) for the standard AUC, (0.238, 0.614) for $pAUC_{0.1}$, and (0.016, 0.272) for IDI. The resulting point estimates and CIs are almost the same as the results with the bandwidths selected via a 10-fold cross validation procedure. In addition, for calculating the weights $\widehat{\omega}_i$, the survival function $G(\cdot)$ of the censoring time C is estimated by a Kaplan-Meier estimator since in the study C is likely to be independent of both T and X, Z. In Sect. 2.1, we commented that if this independence assumption does not hold, we could still provide a correct estimate of $G(\cdot)$ via a semi-parametric model, for example a Cox PH model. Here, we also obtained the estimates of $G(t_0)$ via a Cox PH model, i.e., $\exp\{-\Lambda_0(t_0)\exp(\widehat{\gamma}_c W_c)\}$ where W_c consists of the FRS and the CRP. Based on the resulting weights $\widehat{\omega}_i$, we obtained the point estimates and CIs for the subgroup-specific IncV wrt AUC, $pAUC_{0.1}$ and IDI, which is presented in Fig. 4. The results are very similar to the results using Kaplan–Meier estimator of $G(\cdot)$, and therefore it implies that the independence assumption about the censoring time C is reasonable.

We are also interested in testing whether the subgroup-specific IncV of the CRP over the FRS is constant over the values [0,0.4] of the risk estimates obtained from the FRS. The p

values of testing for constant subgroup-specific IncV are 0.028 for AUC, 0.108 for $pAUC_{0.1}$ and 0.002 for IDI. These results agree with Fig. 5, which shows the point estimates and simultaneous 95% CIs for the relative subgroup-specific IncV wrt AUC, $pAUC_{0.1}$ and IDI. It shows that the subgroup-specific IncVs wrt AUC and IDI are not constant over the interval [0,0.4]; on the other hand,the subgroup-specific IncV wrt $pAUC_{0.1}$ is constant over this interval. It is worth to note that the asymptotic variance of $\widehat{\mathbb{D}}_{\mathbb{A}_{[s_l,s_u]}}(s)$ is larger than that of $\widehat{\mathbb{A}}(s)$, and therefore, the power of testing whether the subgroup-specific IncV is constant over a certain interval is not as strong as the power of testing whether the subgroup-specific IncV is above zero over the interval.

5 Concluding remarks

In this paper, we propose a non-parametric procedure to estimate the IncVs of new markers in prediction accuracy accross different subgroups defined by the conventional scoring system. We also provide the pointwise and simultaneous interval estimates via perturbation resampling. In addition, with proper adjustment for multiple subgroups comparison, our approach is able to systematically identify the subgroups which would benefit from adding new markers. Unlike global measures which do not provide information on how the IncV may vary across subgroups, our methods enables the identification of subgroups for which the new markers may or may not be useful. Existing procedures often assess subgroup-specific IncVs empirically. We provide more rigorous and systematic analytical tools to ensure the validity of such claims and more precisely pinpoint such specific subgroups.

Appropriate choice of prediction accuracy summaries is of great importance to capture the usefulness of new markers. It is also motived by primary research interests. Discrimination is one of the major components in assessing the accuracy of prediction models. The AUC is the most popular summary index which depicts inherent discrimination capacity. However, it is unable to capture how well the predicted risks agree with the actual observed risks (Gail and Pfeiffer 2005). In some cases, alternative summary measures should be also considered, for example, NRI, PCF and PNF. Our approach can be naturally extended to other metrics that maybe more appropriate for particular clinical applications.

The subgroup-specific TPR $\mathscr{S}_1(c;s,t)=pr\left\{\bar{p}_2(X,Z)\geq c|T^{\dagger}\leq t,\bar{p}_1(X)=s\right\}$ and the subgroup-specific FPR $\mathscr{S}_0(c;s,t)=pr\left\{\bar{p}_2(X,Z)\geq c|T^{\dagger}>t,\bar{p}_1(X)=s\right\}$ both depend on the time point *t*, which is usually pre-determined. In some applications, new biomarkers might produce relatively better long-term performance in prediction accuracy than short-term. It is straightforward to extend our procedure to different time points over an arbitrary time interval since the non-parametric estimates of the TPR and FPR, $\widehat{\mathscr{S}}(c;s,t)$, converge to a Gaussian process in time *t*. We could estimate the overall improvement of new markers over a certain time interval by integrating the subgroup-specific pAUC and the subgroup-specific IDI index wrt time *t*. Furthermore, with properly adjusting for multiple comparison, it is possible to identify the time interval where new markers have the most IncVs for different subgroups.

Instead of focusing on the prediction of *t*-year survival for a fixed time point, we might be also interested in a global assessment of a fitted prediction model for the continuous event time. One example of such global measure is the C-statistic of the prediction score $\mathscr{P}_2(X,Z)$, $pr\left\{\bar{p}_2(X_i,Z_i)>\bar{p}_2(X_{i'},Z_{i'})|T_{i'}^{\dagger}>T_i^{\dagger}\right\}$ (Harrell Jr et al. 1996; Korn and Simon 1990;

Pencina and D'Agostino 2004). When the event time $T^\dagger$ is subject to right censoring which may have finite support [0, τ], one may consider a truncated C-statistic,

$$C_\tau = pr\left\{\bar{p}_2(X_i, Z_i) > \bar{p}_2(X_{i'}, Z_{i'}) \,|\, T_{i'}^\dagger > T_i^\dagger, T_i^\dagger < \tau\right\},$$

as considered in Heagerty and Zheng (2005) and Uno et al. (2011a). It is straightforward to extend C_τ to our subgroup-specific C-statistic

$$C_\tau(s) = pr\left\{\bar{p}_2(X_i, Z_i) > \bar{p}_2(X_{i'}, Z_{i'}) \,|\, T_{i'}^\dagger > T_i^\dagger, T_i^\dagger < \tau, \bar{p}_1(X_i) = \bar{p}_1(X_{i'}) = s\right\}$$

and construct an IPW kernel estimator for $C_\tau(s)$ as for other accuracy measures.

Acknowledgments

The Framingham Heart Study and the Framingham SHARe project are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with Boston University. The Framingham SHARe data used for the analyses described in this manuscript were obtained through dbGaP (access number: phs000007.v3.p2). This manuscript was not prepared in collaboration with investigators of the Framingham Heart Study and does not necessarily reflect the opinions or views of the Framingham Heart Study, Boston University, or the NHLBI. The work is supported by Grants U01-CA86368, P01-CA053996, R01-GM085047, R01-GM079330, R01-AI052817 and U54-LM008748 awarded by the National Institutes of Health.

Appendix A

Let $\mathbb{P}_n$ and $\mathbb{P}$ denote expectation with respect to (wrt) the empirical probability measure of $\{(T_i, \Delta_i, X_i, Z_i), i = 1, \ldots, n\}$ and the probability measure of (T, Δ, X, Z), respectively, and $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - \mathbb{P})$. We use $\dot{\mathscr{F}}(x)$ to denote $d\mathscr{F}(x)/dx$ for any function $\mathscr{F}$, $\simeq$ to denote equivalence up to $o_p(1)$, and $\lesssim$ to denote being bounded above up to a universal constant. Let β_0 and γ_0 denote the solution to $E\left[V_i\left\{Y_i^\dagger - g_1\left(\beta' V_i\right)\right\}\right] = 0$ and $E\left[W_i\left\{Y_i^\dagger - g_2\left(\gamma' W_i\right)\right\}\right] = 0$, respectively. Let $\bar{p}_{1i} = g_1\left(\beta_0' V_i\right)$ and $\bar{p}_{2i} = g_2\left(\gamma_0' W_i\right)$. Let $\omega = \Delta I(T \le t_0)/G_{X,Z}(T) + I(T > t_0)/G_{X,Z}(t_0)$, $\widehat{M}_i(c) = I(\widehat{p}_{2i} \ge c)$ and $\bar{M}_i(c) = I\left(\bar{p}_{2i} \ge c\right)$. For $y = 0, 1$, let $f_y(c; s)$ denote the conditional density of $\bar{p}_{2i}$ given $Y_i^\dagger = y$ and $\bar{p}_{1i} = s$ and we assumed that $f_y(c; s)$ is continuous and bounded away from zero uniformly in c and s. This assumption implies that ROC$(u; s)$ has continuous and bounded derivative RÒC$(u; s) = \partial$ROC$(u; s)/\partial u$.

We assume that V and W are bounded, and $\tau(y; s) = \partial pr\left[\phi\left\{\bar{p}_1(X)\right\} \le s, \quad Y^\dagger = y\right]/\partial s$, is continuously differentiable bounded derivatives and bounded away from zero. Throughout, the bandwidths are assumed to be of order $n^{-\nu}$ with $\nu \in (1/5, 1/2)$. For ease of presentation and without loss of generality, we assume that $h_1 = h_0$, denoted by h, and suppress h from the notations. Without loss of generality, we assume that $\sup_{t,x,z}|n^{\frac{1}{2}}\left\{\widehat{G}_{X,Z}(t) - G_{X,Z}(t)\right\}| = O_p(1)$. When C is assumed to be independent of both T and (X, Z), the simple Kaplan–Meier estimator satisfies this condition. When C depends on (X, Z), $\hat{G}_{X,Z}$ obtained under the Cox model also satisfies this condition provided that W_C is bounded. The kernel function K is assumed to be symmetric, smooth with a bounded support on $[-1, 1]$ and we let $m_2 = \int K(x)^2 dx$.

A.1 Asymptotic expansions for $\widehat{\mathscr{S}_y}(c;s)$

Uniform convergence rate for $\widehat{\mathscr{S}_y}(c;s)$ We first establish the following uniform convergence rate of $\widehat{\mathscr{S}_y}(c;s)=g\left\{\widehat{a}_y(c;s)\right\}$:

$$\sup_{s\in\mathscr{I}_h,c}|\widehat{\mathscr{S}_y}(c;s)-S_y(c;s)|=O_p\left\{(nh)^{-\frac{1}{2}}\log(n)\right\}=o_p(1). \quad (6)$$

To this end, we note that for any given c and s,

$$\widehat{\zeta}_y(c;s)=\begin{bmatrix}\widehat{\zeta}_{a_y}(c;s)\\ \widehat{\zeta}_{b_y}(c;s)\end{bmatrix}=\begin{bmatrix}\widehat{a}_y(c;s)-a_y(c;s)\\ \widehat{b}_y(c;s)-b_y(c;s)\end{bmatrix}$$

is the solution to the estimating equation $\widehat{\Psi}_y\left(\zeta_y,c,s\right)=0$, where $\boldsymbol{\zeta}_y=(\zeta_{a_y},\zeta_{b_y})'$ and

$$\begin{aligned}\widehat{\Psi}_y\left(\zeta_y;c,s\right)&=\begin{bmatrix}\widehat{\Psi}_{y1}\left(\zeta_y,c,s\right)\\ \widehat{\Psi}_{y2}\left(\zeta_y,c,s\right)\end{bmatrix}\\&=n^{-1}\sum_{i=Y_i=y}\widehat{w}_i\begin{bmatrix}1\\ h^{-1}\widehat{\varepsilon}_{i1}(s)\end{bmatrix}K_h\{\widehat{\varepsilon}_{i1}(s)\}\\&\quad\times\left[\widehat{M}_i(c)-\mathscr{G}\left\{\zeta_y,c,s;\phi(\widehat{p}_{1i}),h\right\}\right],\end{aligned}$$

$a_y(c;s)=g^{-1}\left\{\mathscr{S}_y(c;s)\right\}$, $b_y(c;s)=\partial g^{-1}\left\{\mathscr{S}_y(c;s)\right\}/\partial s$ and $\mathscr{G}\left(\zeta_y,c,s;e,h\right)=g\left[a_y(c;s)+b_y(c;s)\{e-\phi(s)\}+\zeta_{a_y}+\zeta_{b_y}h^{-1}\{e-\phi(s)\}\right]$. We next establish the convergence rate for $\sup_{\zeta_y,c,s}|\widehat{\Psi}_y\left(\zeta_y;c,s\right)-\Psi_y\left(\zeta_y;c,s\right)|$, where

$$\begin{aligned}\widehat{\Psi}_y\left(\zeta_y;c,s\right)&=\begin{bmatrix}\Psi_{y1}\left(\zeta_y,c,s\right)\\ \Psi_{y2}\left(\zeta_y;c,s\right)\end{bmatrix}\\&=\tau(y;s)\begin{bmatrix}\mathscr{S}_y(c;s)-\int K(t)\,g\left\{a_y(c;s)+\zeta_{a_y}+\zeta_{b_y}t\right\}dt\\ -\int tK(t)\,g\left\{a_y(c;s)+\zeta_{a_y}+\zeta_{b_y}t\right\}dt\end{bmatrix}.\end{aligned}$$

We first show that

$$\sup_{s\in\mathscr{I}_h,c}\left|n^{-1}\sum_{i:Y_i=y}\widehat{\omega}K_h\{\widehat{\varepsilon}_{i1}(s)\}\widehat{M}_i(c)-\tau(y;s)\mathscr{S}_y(c;s)\right|$$

and

$$\sup_{\zeta_y,s\in\mathscr{I}_h,c}\left|n^{-1}\sum_{i:Y_i=y}\widehat{\omega}K_h\{\widehat{\varepsilon}_{i1}(s)\}\mathscr{G}\left\{\zeta_y,c,s;\phi(\widehat{p}_{1i}),h\right\}-\tau(y;s)\int K(t)\,g\left\{a_y(c;s)+\zeta_{a_y}+\zeta_{b_y}t\right\}dt\right|$$

are both $O_p\{(nh)^{-1/2}\log(n)\}$ where $\mathcal{I}_h=\left[\phi^{-1}(\rho_l+h),\phi^{-1}(\rho_u-h)\right]$ and $[\rho_l,\rho_u]$ is a subset of the support of $\phi\left\{g_1\left(\beta_0^T V\right)\right\}$. To this end, we note that since $\sup_u|\hat{G}_{X,Z}(u)-G_{X,Z}(u)|=O_p(n^{-1/2}$ and $|\widehat{\beta}-\beta_0|=O_p\left(n^{-\frac{1}{2}}\right)$,

$$\begin{aligned}&\left|n^{-1}\sum_{i:Y_i=y}(\widehat{\omega}_i-\omega_i)K_h\{\widehat{\varepsilon}_{i1}(s)\}\mathcal{G}\left\{\zeta_y,c,s;\phi(\widehat{p}_{1i}),h\right\}\right|\\&\le n^{-1}\sum_{i:Y_i=y}|\widehat{\omega}_i-\omega_i|K_h\{\widehat{\varepsilon}_{i1}(s)\}\\&=O_p\left(n^{-\frac{1}{2}}\right).\end{aligned}$$

This implies that

$$\begin{aligned}&\left|n^{-1}\sum_{i:Y_i=y}\widehat{\omega}_iK_h\{\widehat{\varepsilon}_{i1}(s)\}\mathcal{G}\left\{\zeta_y,c,s;\phi(\widehat{p}_{1i}),h\right\}-\tau(y;s)\int K(t)\,g\left\{a_y(c;s)+\zeta_{a_y}+\zeta_{b_y}t\right\}dt\right|\\&\le\left|n^{-\frac{1}{2}}\int K_h\{e-\phi(s)\}\mathcal{G}\left\{\zeta_y,c,s;\phi(\widehat{p}_{1i}),h\right\}\times d\mathbb{G}_n\left[\omega I\{\phi(\widehat{p}_{i1})\le e\}-\omega I\left\{\phi\left(\bar{p}_{i1}\right)\le e\right\}\right]\right|\\&+\left|\int K_h\{e-\phi(s)\}\mathcal{G}\left\{\zeta_y,c,s;\phi(\widehat{p}_{1i}),h\right\}d\mathbb{P}\left[\omega I\left\{\phi\left(\bar{p}_{i1}\right)\le e\right\}\right]-\tau(y;s)\int K(t)\,g\left\{a_y(c;s)+\zeta_{a_y}+\zeta_{b_y}t\right\}dt\right|\\&+\left|n^{-\frac{1}{2}}\int K_h\{e-\phi(s)\}d\mathbb{P}\left[\omega\mathcal{G}\left\{\zeta_y,c,s;\phi(\widehat{p}_{1i}),h\right\}I\left\{\phi\left(\bar{p}_{i1}\right)\le e\right\}\right]\right|\\&+O_p\left(n^{-\frac{1}{2}}\right)\lesssim n^{-\frac{1}{2}}h^{-1}\|\mathbb{G}_n\|_{\mathcal{H}_\delta}+\left|n^{-\frac{1}{2}}\int K_h\{e-\phi(s)\}\right.\\&\left.\times d\mathbb{P}\left[\omega\mathcal{G}\left\{\zeta_y,c,s;\phi(\widehat{p}_{1i}),h\right\}I\left\{\phi\left(\bar{p}_{i1}\right)\le e\right\}\right]\right|+O_p\left(n^{-\frac{1}{2}}+h^2\right),\end{aligned}$$

where $\mathcal{H}_\delta=\left\{\omega I\left[\phi\left\{g_1\left(\beta'\nu\right)\right\}\le e\right]-\omega I\left[\phi\left\{g_1\left(\beta_0'\nu\right)\right\}\le e\right]:|\beta-\beta_0|\le\delta,e\right\}$ is a class of functions indexed by β and e. By the maximum inequality of Van der Vaart and Wellner (1996), we have

$$E\|\mathbb{G}_n\|_{\mathcal{H}_\delta}\lesssim\delta^{\frac{1}{2}}\{|\log(\delta)|+|\log(h)|\}\left[1+\frac{\delta^{\frac{1}{2}}\{|\log(\delta)|+|\log(h)|\}}{\delta n^{\frac{1}{2}}}\right]$$

Together with the fact that $|\widehat{\beta}-\beta_0|=O_p\left(n^{-\frac{1}{2}}\right)$ from Uno et al. (2007), it implies that $n^{-\frac{1}{2}}h^{-1}\|\mathbb{G}_n\|_{\mathcal{H}_\delta}=O_p\left\{(nh)^{-\frac{1}{2}}\left(nh^2\right)^{-\frac{1}{4}}\log(n)\right\}$. In addition, with the standard arguments used in Bickel and Rosenblatt (1973), it can be shown that

$$\begin{aligned}&\left|n^{-\frac{1}{2}}\int K_h\{e-\phi(s)\}d\mathbb{P}\left[\omega\mathcal{G}\left\{\zeta_y,c,s;\phi(\widehat{p}_{1i}),h\right\}I\left\{\phi\left(\bar{p}_{i1}\right)\le e\right\}\right]\right|\\&=O_p\left\{(nh)^{-\frac{1}{2}}\log(n)\right\}.\end{aligned}$$

Therefore, for $h=n^{-\nu}$, $1/5<\nu<1/2$,

$$\sup_{\zeta_y, s\in\mathscr{I}_h, c}\left|n^{-1}\sum_{i:Y_i=y}\widehat{\omega}_i K_h\{\widehat{\varepsilon}_{i1}(s)\}\mathscr{G}\left\{\zeta_y, c, s;\phi(\widehat{p}_{1i}), h\right\} - \tau(y;s)\int K(t)\, g\left\{a_y(c;s)+\zeta_{a_y}+\zeta_{b_y}t\right\}dt\right|$$

is $O_p\{(nh)^{-1/2}\log(n)\}$. Following with similar arguments as given above, coupled with the fact that $|\widehat{\gamma} - \gamma_0| = O_p\left(n^{-\frac{1}{2}}\right)$, we have

$$\sup_{s\in\mathscr{I}_h, c\in[0,1]}\left|n^{-1}\sum_{i:Y_i=y}\widehat{\omega}_i K_h\{\widehat{\varepsilon}_{i1}(s)\}\widehat{M}_i(c) - \tau(y;s)\mathscr{S}_y(c;s)\right| = O_p\left\{(nh)^{-\frac{1}{2}}\log(n)\right\}.$$

Thus, $\sup_{\zeta_y,c,s}|\widehat{\Psi}_{y1}\left(\zeta_y;c,s\right) - \Psi_{y1}\left(\zeta_y;c,s\right)| = O_p\left\{(nh)^{-\frac{1}{2}}\log(n)\right\} = o_p(1)$. It follows from same arguments as given above that

$$\sup_{\zeta_y,c,s}|\widehat{\Psi}_{y2}\left(\zeta_y;c,s\right) - \Psi_{y2}\left(\zeta_y;c,s\right)| = O_p\left\{(nh)^{-\frac{1}{2}}\log(n)+h\right\} = o_p(1).$$

Therefore, $\sup_{\zeta_y,c,s}|\widehat{\Psi}_y\left(\zeta_y;c,s\right) - \Psi_y\left(\zeta_y;c,s\right)| = o_p(1)$. In addition, we note that **0** is the unique solution to the equation $\boldsymbol{\psi}_y(\boldsymbol{\zeta}_y; c, s) = 0$ wrt $\boldsymbol{\zeta}_y$. It suggests that $\sup_{s,c}|\widehat{\zeta}_{a_y}(c;s)| = O_p\left\{(nh)^{-\frac{1}{2}}\log(n)\right\} = o_p(1)$, which implies the consistency of $\mathscr{S}_y(c;s)$,

$$\sup_{s\in\mathscr{I}_h, c\in[0,1]}|\widehat{\mathscr{S}_y}(c;s) - \mathscr{S}_y(c;s)| = O_p\left\{(nh)^{-\frac{1}{2}}\log(n)\right\} = o_p(1).$$

Asymptotic expansion for $\widehat{\mathscr{S}_y}(c;s)$ Let $\widehat{d}_y(c;s) = \sqrt{nh}\left\{\widehat{a}_y(c;s) - a_y(c;s)\right\}$. It follows from a Taylor series expansion and the convergence rate of $\boldsymbol{\zeta}_y(c; s)$ that

$$\widehat{d}_y(c;s) = \frac{\sqrt{nh}\mathbb{P}_n\left(\widehat{\omega}I(Y=y)K_h\{\widehat{\varepsilon}_1(s)\}\left[\widehat{M}(c) - \mathscr{G}_y^0\{c, s;\phi(\widehat{p}_1)\}\right]\right)}{\tau\{y;\phi(s)\}\dot{g}\left\{a_y(c;s)\right\}} + o_p(1), \quad (7)$$

where $\mathscr{G}_y^0(c, s;e) = g\left[a_y(c;s) + b_y(c;s)\{e - \phi(s)\}\right]$. Futhermore, since $\sup_{t\le t_0}\{\hat{G}_{X,Z}(t) - G_{X,Z}(t)| = O_p(n^{-1/2})$,

$$\widehat{d}_y(c;s) = \frac{\sqrt{nh}\mathbb{P}_n\left(\omega I(Y=y)K_h\{\widehat{\varepsilon}_1(s)\}\left[\widehat{M}(c) - \mathscr{G}_y^0\{c, s;\phi(\widehat{p}_1)\}\right]\right)}{\tau\{y;\phi(s)\}\dot{g}\left\{a_y(c;s)\right\}} + o_p(1).$$

We next show that $\widehat{d}_y(c;s)$ is asymptotically equivalent to

$$\tilde{d}_y(c;s) = \frac{\sqrt{nh}\mathbb{P}_n\left(\omega I(Y=y) K_h\left\{\bar{\varepsilon}_1(s)\right\}\left[\bar{M}(c) - \mathscr{G}_y^0\left\{c, s; \phi\left(\bar{p}_1\right)\right\}\right]\right)}{\tau\{y;\phi(s)\}\, \dot{g}\left\{a_y(c;s)\right\}}, \quad (8)$$

where $\bar{\varepsilon}_1(s) = \phi\left(\bar{p}_1\right) - \phi(s)$. From (8) and the fact that $\tau\{y; \phi(s)\}$ is bounded away from 0 uniformly in s, we have

$$\begin{aligned}
&|\widehat{d}_y(s) - \tilde{d}_y(s)| \\
&\lesssim h^{\frac{1}{2}}\left|\int K_h\{e - \phi(s)\}\, d\mathbb{G}_n\left(I(Y=y)\,\omega\left[\widehat{M}(c)\, I\{\phi(\widehat{p}_1) \le e\} - \bar{M}(c)\, I\left\{\phi\left(\bar{p}_1\right) \le e\right\}\right]\right)\right| \\
&+h^{\frac{1}{2}}\left|\int K_h\{e - \phi(s)\}\, \mathscr{G}_y(c, s; e)\, d\mathbb{G}_n\left(I(Y=y)\left[\omega I\{\phi(\widehat{p}_1) \le e\} - \omega I\left\{\phi\left(\bar{p}_1\right) \le e\right\}\right]\right)\right| \\
&+\left|\sqrt{nh}\int K_h\{e - \phi(s)\}\, d\mathbb{P}\left(I(Y=y)\left[\omega\widehat{M}(c)\, I\{\phi(\widehat{p}_1) \le e\} - \omega\, \bar{M}(c)\, I\left\{\phi\left(\bar{p}_1\right) \le e\right\}\right]\right)\right| \\
&+\left|\sqrt{nh}\int K_h\{e - \phi(s)\}\, d\mathbb{P}\left(I(Y=y)\left[\omega\mathscr{G}_y\{c, s; \phi(\widehat{p}_1)\}\, I\{\phi(\widehat{p}_1) \le e\} - \omega\mathscr{G}_y\left\{c, s; \phi\left(\bar{p}_1\right)\right\} I\left\{\phi\left(\bar{p}_1\right) \le e\right\}\right]\right)\right| \\
&\lesssim h^{\frac{1}{2}}\|\mathbb{G}_n\|_{\mathscr{F}_\delta} + h^{\frac{1}{2}}\|\mathbb{G}_n\|_{\mathscr{H}_\delta} + O_p\left\{(nh)^{1/2}|\widehat{\beta} - \beta_0| + |\widehat{\gamma} - \gamma_0| + h^2\right\},
\end{aligned}$$

where

$\mathscr{F}_\delta = \left\{\omega I\left\{g_2\left(\gamma'\omega\right) \ge c\right\} I\left[\phi\left\{g_1\left(\beta'\nu\right)\right\} \le e\right] - \omega I\left\{g_2\left(\gamma_0'\omega\right) \ge c\right\} I\left[\phi\left\{g_1\left(\beta_0'\nu\right)\right\} \le e\right] : |\gamma - \gamma_0| + |\beta - \beta_0| \le \delta, e\right\}$ is the class of functions indexed by γ, β and e. By the maximum inequality of Van der Vaart and Wellner (1996) and the fact that $|\widehat{\beta} - \beta_0| + |\widehat{\gamma} - \gamma_0| = O_p\left(n^{-\frac{1}{2}}\right)$ from Uno et al. (2007), we have $h^{\frac{1}{2}}\|\mathbb{G}_n\|_{\mathscr{F}_\delta} = O_p\left\{h^{-\frac{1}{2}} n^{-\frac{1}{4}} \log(n)\right\}$ and $h^{\frac{1}{2}}\|\mathbb{G}_n\|_{\mathscr{H}_\delta} = O_p\left\{h^{-\frac{1}{2}} n^{-\frac{1}{4}} \log(n)\right\}$. It follows that $\sup_s|\widehat{d}_y(s) - \tilde{d}_y(s)| = o_p(1)$. Then, by a delta method,

$$\widehat{\mathscr{W}}_{\mathscr{S}_y}(c;s) = \sqrt{nh}\left\{\widehat{\mathscr{S}}_y(c;s) - \mathscr{S}_y(c;s)\right\} \simeq \sqrt{nh}\mathbb{P}_n\left[K_h\left\{\bar{\varepsilon}_1(s)\right\} \mathscr{D}_{\mathscr{S}_y}(c;s)\right] \quad (9)$$

where

$$\mathscr{D}_{\mathscr{S}_y}(c;s) = \tau\{y;\phi(s)\}^{-1}\omega I(Y=y)\left\{\bar{M}(c) - \mathscr{S}_y(c;s)\right\} \quad (10)$$

Using the same arguments as for establishing the uniform convergence rate of conditional Kaplan-Meier estimators (Dabrowska 1989; Du and Akritas 2002), we obtain (6). Furthermore, following similar arguments as given in Dabrowska (1987, 1997), we have $\widehat{\mathscr{W}}_{\mathscr{S}_y}(c;s)$ converges weakly to a Gaussian process in c for all s. Note that as for all kernel estimators, $\widehat{\mathscr{W}}_{\mathscr{S}_y}(c;s)$ does not converge as a process in s.

A.2 Uniform consistency of $\widehat{pAUC}_f(s)$

Next we establish the uniform convergence rate for $\widehat{\mathrm{ROC}}(u;s)$. To this end, we write

$$\widehat{\mathrm{ROC}}(u;s) - \mathrm{ROC}(u;s) = \widehat{\varepsilon}_1(u;s) + \widehat{\varepsilon}_0(u;s),$$

where $\widehat{\varepsilon}(u;s)=\widehat{\mathscr{S}_1}\left\{\widehat{\mathscr{S}_0^{-1}}(u;s);s\right\}-\mathscr{S}_1\left\{\widehat{\mathscr{S}_0^{-1}}(u;s);s\right\}$ and $\widehat{\varepsilon}(u;s)=\mathscr{S}_1\left\{\widehat{\mathscr{S}_0^{-1}}(u;s);s\right\}-\mathscr{S}_1\left\{\mathscr{S}_0^{-1}(u;s);s\right\}$. It follows from (6) that $\sup_{u;s}|\widehat{\varepsilon}_1(u;s)|\le\sup_{c;s}|\widehat{\mathscr{S}_1}(c;s)-\mathscr{S}_1(c;s)|$. Let $\widehat{\mathscr{I}}(u;s)=\mathscr{S}_0\left\{\widehat{\mathscr{S}_0^{-1}}(u;s);s\right\}$. Then $\widehat{\varepsilon}_0(u;s)=\mathrm{ROC}\{\mathscr{I}(u;s);s\}-\mathrm{ROC}(u;s)$. Noting that $\sup_u|\widehat{\mathscr{I}}(u;s)-u|=\sup_u|\widehat{\mathscr{I}}(u;s)-\widehat{\mathscr{S}_0}\left\{\widehat{\mathscr{S}_0^{-1}}(u;s);s\right\}|+n^{-1}\le\sup_c|\mathscr{S}_0(c;s)-\widehat{\mathscr{S}_0}(c;s)|+n^{-1}=O_p\left\{(nh)^{-1/2}\log n\right\}$, we have $\widehat{\varepsilon}_0(u;s)=O_p\left\{(nh)^{-1/2}\log n\right\}$ by the continuity and boundedness of $\dot{\mathrm{R}}\mathrm{OC}(u,s)$. Therefore,

$$\sup_{u,s}|\widehat{\mathrm{ROC}}(u;s)-\mathrm{ROC}(u;s)|=O_p\left\{(nh)^{-1/2}\log n\right\}$$

which implies

$$\begin{aligned}\sup_{s\in\mathscr{I}_h}&\left|\widehat{\mathrm{pAUC}}_f(s)-\mathrm{pAUC}_f(s)\right|\\&\lesssim\sup_{s\in\mathscr{I}_h}\int_0^f\left|\widehat{\mathrm{ROC}}(u;s)-\mathrm{ROC}(u;s)\right|du=O_p\left\{(nh)^{-\frac{1}{2}}\log n\right\}.\end{aligned}$$

and hence the uniform consistency of $\widehat{\mathrm{pAUC}}_f(s)$.

A.3 Asymptotic distribution of $\widehat{\mathscr{W}}_{\mathrm{pAUC}_f}(s)$

To derive the asymptotic distribution for $\widehat{\mathscr{W}}_{\mathrm{pAUC}_f}(s)$, we first derive asymptotic expansions for $\widehat{\mathscr{W}}_{\mathrm{ROC}}(u;s)=\sqrt{nh}\left\{\widehat{\mathrm{ROC}}(u;s)-\mathrm{ROC}(u;s)\right\}=\sqrt{nh}\quad\widehat{\varepsilon}_1(u;s)+\sqrt{nh}\quad\widehat{\varepsilon}_0(u;s)$. From the weak convergence of $\widehat{\mathscr{W}}_{\mathscr{S}_y}(c;s)$ in c, the approximation in (9), and the consistency of $\widehat{\mathscr{S}_0^{-1}}(c;s)$ given in the Appendix A.2, we have

$$\begin{aligned}\sqrt{nh}\widehat{\varepsilon}_1(u;s)&\simeq\sqrt{nh}\left[\widehat{\mathscr{S}_1}\left\{\mathscr{S}_0^{-1}(u;s);s\right\}-\mathrm{ROC}(u;s)\right]\\&\simeq\sqrt{nh}\mathbb{P}_n\left[K_h\left\{\bar{\varepsilon}_1(s)\right\}\mathscr{D}_{\mathscr{S}_1}\left\{\mathscr{S}_0^{-1}(u;s);s\right\}\right]\end{aligned}$$

On the other hand, from the uniform convergence of $\widehat{\mathscr{I}_0}(u;s)\to u$ and the weak convergence of $\widehat{\mathscr{D}_0}(c;s)$ in c, we have

$$\begin{aligned}\sqrt{nh}\left\{u-\widehat{\mathscr{I}}(u;s)\right\}&\simeq\sqrt{nh}\left[\widehat{\mathscr{I}^{-1}}\left\{\widehat{\mathscr{I}}(u;s);s\right\}-\widehat{\mathscr{I}}(u;s)\right]\simeq\sqrt{nh}\left\{\widehat{\mathscr{I}^{-1}}(u;s)-u\right\}\\&\simeq\sqrt{nh}\left[\widehat{\mathscr{S}_0}\left\{\mathscr{S}_0^{-1}(u;s);s\right\}-u\right]\end{aligned}$$

This, together with a Taylor series expansion and the expansion given (9), implies that

$$\sqrt{nh}\widehat{\varepsilon}_0(u;s)\simeq-R\ \dot{O}\ C(u;s)\,\mathbb{P}_n\left[K_h\left\{\bar{\varepsilon}_1(s)\right\}\mathscr{D}_{\mathscr{S}_0}\left\{\mathscr{S}_0^{-1}(u;s);s\right\}\right]$$

It follows that

$$\widehat{\mathscr{W}}_{\mathrm{pAUC}f}(s) \simeq \sqrt{nh}\mathbb{P}_n\left[K_h\left\{\bar{\varepsilon}_1(s)\right\}\mathscr{D}_{\mathrm{pAUC}f}(s)\right] \quad (11)$$

where

$$\mathscr{D}_{\mathrm{pAUC}f}(s)=\int_0^f\left[\mathscr{D}_{\mathscr{S}_1}\left\{\mathscr{S}_0^{-1}(u;s);s\right\}-R\,\dot{O}\,C(u;s)\,\mathscr{D}_{\mathscr{S}_0}\left\{\mathscr{S}_0^{-1}(u;s);s\right\}\right]du. \quad (12)$$

It then follows from a central limit theorem that for any fixed s, $\widehat{\mathscr{W}}_{\mathrm{pAUC}_f}(s)$ converges to a normal with mean 0 and variance

$$\sigma^2_{\mathrm{pAUC}_f}(s)=m_2\left[\tau\{1;\phi(s)\}\dot{F}_{\phi(\bar{p}_1)}(s)\right]^{-1}\sigma_1^2(s)+m_2\left[\tau\{0;\phi(s)\}\dot{F}_{\phi(\bar{p}_1)}(s)\right]^{-1}\sigma_0^2(s),$$

where $\dot{F}_{\phi(\bar{p}_1)}(s)$ is the density function of $\phi(\bar{p}_1)$,

$$\sigma_1^2(s)=E\left(G(T^\dagger)^{-1}\left[\int_0^f\bar{M}\left\{\mathscr{S}_0^{-1}(u;s)\right\}du-\mathrm{pAUC}_f(s)\right]^2\middle|\bar{p}_1=s,\quad Y^\dagger=1\right),$$

and

$$\sigma_0^2(s)=E\left(G(t_0)^{-1}\left[\int_0^f\bar{M}\left\{\mathscr{S}_0^{-1}(u;s)\right\}d\mathrm{ROC}(u;s)-\int_0^f u d\mathrm{ROC}(u;s)\right]^2\middle|\bar{p}_1=s,\quad Y^\dagger=0\right).$$

A.4 Justification for the resampling methods

To justify the resampling method, we first note that $|\beta^*-\widehat{\beta}|+|\gamma^*-\widehat{\gamma}|+\sup_{t\le t_0}|G^*_{X,Z}(t)-\widehat{G}_{X,Z}(t)|=O_p\left(n^{-\frac{1}{2}}\right)$. It follows from similar arguments given in the Appendix A and Appendix 1 of Cai et al. (2010) that $\mathscr{W}^*_{\mathscr{S}_y}(c;s)=\sqrt{nh}\left\{\mathscr{S}^*_y(c;s)-\widehat{\mathscr{S}_y}(c;s)\right\}\simeq n^{\frac{1}{2}}h^{-1/2}\sum_{i=1}^n\widehat{\mathscr{D}}_{\mathscr{S}_{yi}}(c;s)\xi_i$, where $\widehat{\mathscr{D}}_{\mathscr{S}_{yi}}(c;s)$ is obtained by replacing all theoretical quantities in $\mathscr{D}_{\mathscr{S}_y}(c;s)$ given in (10) with the estimated counterparts for the ith subject. This, together with similar arguments as given above for the expansion of $\widehat{\mathscr{W}}_{\mathrm{ROC}}(u;s)$, implies that

$$\begin{aligned}\mathscr{W}^*_{\mathrm{pAUC}f}(s) &= \int_0^f\sqrt{nh}\left\{\mathrm{ROC}^*(u;s)-\widehat{\mathrm{ROC}}(u;s)\right\}du\\ &\simeq n^{-\frac{1}{2}}h^{-1/2}\sum_{i=1}^n K_h\{\widehat{\varepsilon}_1(s)\}\,\widehat{\mathscr{D}}_{\mathrm{pAUC}f}(s)\xi_i,\end{aligned}$$

where $\widehat{\mathscr{D}}_{\mathrm{pAUC}f}(s)=\int_0^f\left[\widehat{\mathscr{D}}_{\mathscr{S}_1 i}\left\{\widehat{\mathscr{S}_0}^{-1}(u;s);s\right\}-R\,\dot{O}\,C(u;s)\,\widehat{\mathscr{D}}_{\mathscr{S}_0 i}\left\{\widehat{\mathscr{S}_0}^{-1}(u;s);s\right\}\right]du$ Conditional on the data, $\mathscr{W}^*_{\mathrm{pAUC}_f}(s)$ is approximately normally distributed with mean 0 and variance

$$\widehat{\sigma}^2_{\text{pAUC}_f}(s) = h^{-1}\sum_{i=1}^{n} K_h\{\widehat{\varepsilon}(s)\}^2 \widehat{\mathscr{D}}_{\text{pAUC}\,f}(s)^2.$$

Using the consistency of the proposed estimators along with similar arguments as given above, it is not difficult to show that the above variance converges to $\sigma^2_{\text{pAUC}_f}(s)$ as $n \to \infty$. Therefore, the empirical distribution obtained from the perturbed sample can be used to approximate the distribution of $\widehat{\mathscr{W}}_{\text{pAUC}_f}(s)$.

We now show that after proper standardization, the supermum type statistics Γ converges weakly. To this end, we first note that, similar arguments as given in the Appendix A can be used to show that $\sup_{s\in\mathscr{I}_h} |\widehat{\sigma}^2_{\text{pAUC}_f}(s) - \sigma^2_{\text{pAUC}_f}(s)| = o_p\left(n^{-\delta}\right)$ and

$$\Gamma = \sup_{s\in\mathscr{I}_h} \left| \frac{\sqrt{nh}\mathbb{P}_n\left[K_h\left\{\bar{\varepsilon}_1(s)\right\}\mathscr{D}_{\text{pAUC}\,f}(s)\right]}{\sigma_{\text{pAUC}\,f}(s)} \right| + o_p\left(n^{-\delta}\right),$$

for some small positive constant δ. Using similar arguments in Bickel and Rosenblatt (1973), we have

$$pr\{a_n(\Gamma - d_n) < x\} \to e^{-2e^{-x}},$$

where $a_n = [2\log\{(\rho_u - \rho_l)/h\}]^{1/2}$ and $d_n = a_n + a_n^{-1}\log\left\{\int \dot{K}(t)^2 dt/(4m_2\pi)\right\}$. Now justify the resampling procedure for constructing the CI, we note that

$$\mathscr{W}^*_{\text{pAUC}\,f}(s) = n^{-\frac{1}{2}}h^{\frac{1}{2}}\sum_{i=1}^{n} K_h\{\widehat{\varepsilon}_{1i}(s)\}\widehat{\mathscr{D}}_{\text{pAUC}\,f}(s)(\xi_i - 1) + \varepsilon^*(s)$$

where $pr\{\sup_{s\in\Omega(h)}|n^{\varepsilon}\varepsilon^*(s)| \geq e \mid \text{data}\} \to 0$ in probability. Therefore,

$$\Gamma^* = \sup_{s\in\mathscr{I}_h} \left| \frac{n^{-\frac{1}{2}}h^{\frac{1}{2}}\sum_{i=1}^{n} K_h\{\widehat{\varepsilon}_{1i}(s)\}\widehat{\mathscr{D}}_{\text{pAUC}\,f}(s)(\xi_i - 1)}{\sigma_{\text{pAUC}\,f}(s)} \right| + \left|\varepsilon^*_{\sup}\right|.$$

where $pr\left\{|n^{\delta}\varepsilon^*_{\sup}| \geq e|\text{data}\right\} \to 0$. It follows from similar arguments as given in Tian et al. (2005) and Zhao et al. (2010) that

$$\sup\left|pr\{a_n(\Gamma^* - d_n) < x|\text{data}\} - e^{-2e^{-x}}\right| \to 0,$$

in probability as $n \to \infty$. Thus, the conditional distribution of $a_n(\Gamma^* - d_n)$ can be used to approximate the unconditional distribution of $a_n(\Gamma - d_n)$. When $h_0 = h_1$, in general, the standardized Γ does not converge to the extreme value distribution. However, when $h_0 = h_1$

= $k \in (0, \infty)$, the distribution of the suitable standardized version of Γ still can be approximated by that of the corresponding standardized Γ^* conditional on the data (Gilbert et al. 2002).

Appendix B

B.1 Bandwidth selection for $pAUC_f(s)$

The choice of the bandwidths h_0 and h_1 is important for making inference about $\mathscr{S}_y(c;s)$ and consequently $pAUC_f(s)$. Here we propose a two-stage K-fold cross-validation procedure to obtain the optimal bandwidth for $\widehat{\mathscr{S}^{-1}_{0,h_0}}(u;s)$ and $\widehat{\mathscr{S}_{1,h_1}}(c;s)$ sequentially. Specifically, we randomly split the data into K disjoint subsets of about equal sizes denoted by $\{\mathscr{J}_k, k=1,\ldots,K\}$. The two-stage procedure is described as follows:

I. Motivated by the fact that $\mathscr{S}_0^{-1}(u;s)$ is essentially the $(1-u)$-th quantile of the conditional distribution of $\bar{p}_2(X,Z)$ given $Y^\dagger = 0$ and $\bar{p}_1(X) = s$, for each k, we use all the observations not in $\mathscr{J}_k$ to estimate $q_{0,1-u}(s) = \mathscr{S}_0^{-1}(u;s)$ by obtaining $\{\widehat{\alpha}_0(s;h), \widehat{\alpha}_1(s;h)\}$, the minimizer of

$$\sum_{j\in\mathscr{J}_l, l\neq k} I\left(Y_j=0\right)\widehat{w}_j K_h\left\{\widehat{\varepsilon}_{1j}(s)\right\}\rho_{1-u}\left[\widehat{p}_{2j} - g\left\{\alpha_0+\alpha_1\widehat{\varepsilon}_{1j}(s)\right\}\right]$$

wrt (α_0, α_1), where $\rho_\tau(e)$ is a check function defined as $\rho_\tau(e) = \tau e$, if $e \geq 0$; $= (\tau - 1)e$, otherwise. Let $\widehat{q}^{(-k)}_{0,1-u}(s;h) = g\{\widehat{\alpha}_0(s;h)\}$ denote the resulting estimator of $q_{0,1-u}(s) = \mathscr{S}_0^{-1}(u;s)$. With observation in $\mathscr{J}_k$, we obtain

$$Err_k^{(q0)}(h) = \sum_{i\in\mathscr{J}k}(1-Y_i)\widehat{w}_i\int_0^f \rho_{1-u}\left[\widehat{p}_{2i} - \widehat{q}^{(-k)}_{0,1-u}(\widehat{p}_{1i};h)\right]du.$$

Then, we let $h_0^{\text{opt}} = \arg \min_h \sum_{k=1}^K Err_k^{(q_0)}(h)$.

II. Next, to find an optimal h_1 for $\widehat{\mathscr{S}_{1,h_1}}(\cdot;s)$, we choose an error function that directly relates to $\text{pAUC}_f(s) = -\int_{\mathscr{S}_0^{-1}(f;s)}^{\infty}\mathscr{S}_1(c;s)\,d\mathscr{S}_0(c;s)$. Specifically, noting the fact that

$$E\left(\int_{\mathscr{S}_0^{-1}(f;s)}^{\infty}\left[I\left\{g_2\left(\gamma' W_i\right)\geq c\right\} - \mathscr{S}_1(c;s)\right]d\mathscr{S}_0(c;s)\,\middle|\,Y_i^\dagger=1,\quad g_1\left(\beta' X_i\right)=s\right)=0,$$

we use the corresponding mean integrated squared error for $I\left\{g_2\left(\gamma' W_i\right)\geq c\right\} - \mathscr{S}_1(c;s)$ as the error function. For each k, we use all the observations which are not in $\mathscr{J}_k$ to obtain the estimate of $\mathscr{S}_1(c;s)$, denoted by $\widehat{\mathscr{S}}^{(-k)}_{1,h}(c;s)$ via (4). Then, with the observations in $\mathscr{J}_k$, we calculate the prediction error

$$Err_k^{(\mathscr{S}_1)}(h) = -\sum_{i\in\mathscr{J}_k, Y_i=1} \widehat{w}_i \int_{\widehat{\mathscr{S}}_{0,h_0}^{-1}(f;\widehat{p}_{1i})}^{\infty} \left\{ I(\widehat{p}_{2i} \geq c) - \widehat{\mathscr{S}}_{1,h}^{(-k)}(c;\widehat{p}_{1i}) \right\}^2 d\widehat{\mathscr{S}}_{0,h_0}(c;\widehat{p}_{1i}).$$

We let $h_1^{\text{opt}} = \arg \min_h \sum_{k=1}^{K} Err_k^{(\mathscr{S}_1)}(h)$.

Since the order of h_y^{opt} is expected to be $n^{-1/5}$ (Fan and Gijbels 1995), the bandwidth we use for estimation is $h_y = h_y^{\text{opt}} \times n^{-d_0}$ with $0 < d < 3/10$ such that $h_y = n^{-\nu}$ with $1/5 < \nu < 1/2$. This ensures that the resulting functional estimator $\mathscr{S}_{y,h_y}(c;s)$ with the data-dependent smooth parameter has the above desirable large sample properties.

B.2 Bandwidth selection for IDI(s)

Same as bandwidth selection for pAUC, we also propose a K-fold cross validation procedure to choose the optimal bandwidth h_1 for $\text{IS}(s) = \int_0^1 \mathscr{S}_1(c;s)\, dc$ and h_0 for $\text{IP}(s) = \int_0^1 \mathscr{S}_0(c;s)\, dc$ separately. The procedure is described as follows: we randomly split the data into K disjoint subsets of about equal sizes denoted by $\{\mathscr{J}_k, k=1, \ldots, K\}$. Motivated by the fact (3), for each k, we use all the observations not in $\mathscr{J}_k$ to estimate $\int_0^1 \mathscr{S}_y(c, s)\, dc$ by obtaining $\left\{\widehat{\varphi}_0^{(y)}(s;h), \widehat{\varphi}_1^{(y)}(s;h)\right\}$ for $y = 0, 1$, which is the solution to the estimating equation

$$\sum_{j\in\mathscr{J}_l, l\neq k} I\left(Y_j = y\right) \widehat{w}_j K_h \left\{\widehat{\varepsilon}_{1j}(s)\right\} \left[\widehat{p}_{2j} - g\left\{\varphi_0^{(y)} + \varphi_1^{(y)} \widehat{\varepsilon}_{1j}(s)\right\}\right] = 0,$$

wrt $\left(\varphi_0^{(y)}, \varphi_1^{(y)}\right)$. Let $\widehat{\text{IS}}^{(-k)}(s;h) = g\left\{\widehat{\varphi}_0^{(1)}(s;h)\right\}$ and $\widehat{\text{IP}}^{(-k)}(s;h) = g\left\{\widehat{\varphi}_0^{(0)}(s;h)\right\}$. With observations in $\mathscr{J}_k$, we obtain

$$Err_k^{(\text{IS})}(h) = \sum_{i\in\mathscr{J}_k} Y_i \widehat{w}_i \left\{\widehat{p}_{2i} - \widehat{\text{IS}}^{(-k)}(\widehat{p}_{1i};h)\right\}^2,$$

or

$$Err_k^{(\text{IP})}(h) = \sum_{i\in\mathscr{J}_k} (1 - Y_i) \widehat{w}_i \left\{\widehat{p}_{2i} - \widehat{\text{IP}}^{(-k)}(\widehat{p}_{1i};h)\right\}^2,$$

Then, we let $h_1^{opt} = \arg \min_h \sum_{k=1}^{K} Err_k^{(\text{IS})}(h)$ and $h_1^{opt} = \arg \min_h \sum_{k=1}^{K} Err_k^{(\text{IP})}(h)$.

Appendix C

R codes for application will be available from the corresponding author upon request.

References

Baker S, Pinsky P. A proposed design and analysis for comparing digital and analog mammography: special receiver operating characteristic methods for cancer screening. J Am Stat Assoc. 2001; 96:421–428.

Bickel P, Rosenblatt M. On some global measures of the deviations of density function estimates. Ann Stat. 1973; 1:1071–1095.

Blumenthal R, Michos E, Nasir K. Further improvements in CHD risk prediction for women. J Am Med Assoc. 2007; 297:641–643.

Cai T, Cheng S. Robust combination of multiple diagnostic tests for classifying censored event times. Biostatistics. 2008; 9:216–233. [PubMed: 18056687]

Cai T, Dodd LE. Regression analysis for the partial area under the ROC curve. Stat Sin. 2008; 18:817–836.

Cai T, Tian L, Wei L. Semiparametric Box–Cox power transformation models for censored survival observations. Biometrika. 2005; 92(3):619–632.

Cai T, Tian L, Uno H, Solomon S, Wei L. Calibrating parametric subject-specific risk estimation. Biometrika. 2010; 97(2):389–404. [PubMed: 23049123]

Cook N, Ridker P. The use and magnitude of reclassification measures for individual predictors of global cardiovascular risk. Ann Intern Med. 2009; 150(11):795–802. [PubMed: 19487714]

Cook N, Buring J, Ridker P. The effect of including C-reactive protein in cardiovascular risk prediction models for women. Ann Intern Med. 2006; 145:21–29. [PubMed: 16818925]

Cox D. Regression models and life-tables. J R Stat Soc B. 1972; 34(2):187–220.

Dabrowska D. Non-parametric regression with censored survival time data. Scand J Stat. 1987; 14(3): 181–197.

Dabrowska D. Uniform consistency of the kernel conditional Kaplan–Meier estimate. Ann Stat. 1989; 17(3):1157–1167.

Dabrowska D. Smoothed Cox regression. Ann Stat. 1997; 25(4):1510–1540.

D'Agostino R. Risk prediction and finding new independent prognostic factors. J Hypertens. 2006; 24(4):643–645. [PubMed: 16531791]

Dodd L, Pepe M. Partial AUC estimation and regression. Biometrics. 2003; 59:614–623. [PubMed: 14601762]

Du Y, Akritas M. Iid representations of the conditional Kaplan–Meier process for arbitrary distributions. Math Methods Stat. 2002; 11:152–182.

Dwyer AJ. In pursuit of a piece of the ROC. Radiology. 1996; 201:621–625. [PubMed: 8939207]

Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the third report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). J Am Med Assoc. 2001; 285(19):2486–2497.

Fan J, Gijbels I. Data-driven bandwidth selection in local polynomial regression: variable bandwidth selection and spatial adaptation. J R Stat Soc B. 1995; 57:371–394.

Gail M, Pfeiffer R. On criteria for evaluating models of absolute risk. Biostatistics. 2005; 6(2):227–239. [PubMed: 15772102]

Gilbert P, Wei L, Kosorok M, Clemens J. Simultaneous inferences on the contrast of two hazard functions with censored observations. Biometrics. 2002; 58(4):773–780. [PubMed: 12495131]

Harrell F Jr, Lee K, Mark D. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996; 15(4):361–387. [PubMed: 8668867]

Heagerty P, Zheng Y. Survival model predictive accuracy and ROC curves. Biometrics. 2005; 61:92–105. [PubMed: 15737082]

Jiang Y, Metz C, Nishikawa R. A receiver operating characteristic partial area index for highly sensitive diagnostic tests. Radiology. 1996; 201:745–750. [PubMed: 8939225]

Jin Z, Ying Z, Wei L. A simple resampling method by perturbing the minimand. Biometrika. 2001; 88(2):381–390.

Korn E, Simon R. Measures of explained variation for survival data. Stat Med. 1990; 9(5):487–503. [PubMed: 2349402]

An approach to nonparametric regression for life history data using local linear fitting. Ann Stat. 23:787–823.

McIntosh M, Pepe M. Combining several screening tests: optimality of the risk score. Biometrics. 2002; 58(3):657–664. [PubMed: 12230001]

Park Y, Wei L. Estimating subject-specific survival functions under the accelerated failure time model. Biometrika. 2003; 9:717–723.

Park B, Kim W, Ruppert D, Jones M, Signorini D, Kohn R. Simple transformation techniques for improved non-parametric regression. Scand J Stat. 1997; 24(2):145–163.

Paynter N, Chasman D, Pare G, Buring J, Cook N, Miletich J, Ridker P. Association between a literature-based genetic risk score and cardiovascular events in women. J Am Med Assoc. 2010; 303(7):631–637.

Pencina M, D'Agostino R. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Stat Med. 2004; 23(13):2109–2123. [PubMed: 15211606]

Pencina M, D'Agostino RS, D'Agostino RJ, Vasan R. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond (with Coomentaries & Rejoinder). Stat Med. 2008; 27:157–212. [PubMed: 17569110]

Pfeiffer R, Gail M. Two criteria for evaluating risk prediction models. Biometrics. 2010; 67(3):1057–1065. [PubMed: 21155746]

Pfeffer M, Jarcho J. The charisma of subgroups and the subgroups of CHARISMA. N Engl J Med. 2006; 354(16):1744–1746. [PubMed: 16531617]

Ridker P. C-reactive protein and the prediction of cardiovascular events among those at intermediate risk: moving an inflammatory hypothesis toward consensus. J Am Coll Cardiol. 2007; 49(21): 2129–2138. [PubMed: 17531663]

Ridker P, Rifai N, Rose L, Buring J, Cook N. Comparison of C-reactive protein and low-density lipoprotein cholesterol levels in the prediction of first cardiovascular events. N Engl J Med. 2007; 347:1557–1565. [PubMed: 12432042]

Robins J, Ya'Acov R. Toward a curse of dimensionality appropriate (CODA) asymptotic theory for semi-parametric models. Stat Med. 1997; 16(3):285–319. [PubMed: 9004398]

Rothwell P. Treating Individuals 1: External validity of randomised controlled trials: "To whom do the results of this trial apply?". Lancet. 2005; 365:82–93. [PubMed: 15639683]

Tian L, Zucker D, Wei L. On the Cox model with time-varying regression coefficients. J Am Stat Assoc. 2005; 100(469):172–183.

Tian L, Cai T, Wei LJ. Identifying subjects who benefit from additional information for better prediction of the outcome variables. Biometrics. 2009; 65:894–902. [PubMed: 18945268]

Tibshirani R, Hastie T. Local likelihood estimation. J Am Stat Assoc. 1987; 82(398):559–567.

Tice J, Cummings S, Ziv E, Kerlikowske K. Mammographic breast density and the Gail model for breast cancer risk prediction in a screening population. Breast Cancer Res Treat. 2005; 94(2):115–122. [PubMed: 16261410]

Uno H, Cai T, Tian L, Wei L. Evaluating prediction rules for t-year survivors with censored regression models. J Am Stat Assoc. 2007; 102:527–537.

Uno H, Cai T, Pencina M, D'Agostino R, Wei L. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. Stat Med. 2011a; 30(10):1105–1117. [PubMed: 21484848]

Uno H, Cai T, Tian L, Wei LJ. Graphical procedures for evaluating overall and subject-specific incremental values from new predictiors with censored event time data. Biometrics. 2011b; 67:1389–1396. [PubMed: 21504421]

Van der Vaart, AW.; Wellner, JA. Weak convergence and empirical processes. Springer; New York: 1996.

Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson H, Diver W, Thun M, Cox D, Hankinson S, Kraft P, et al. Performance of common genetic variants in breast-cancer risk models. N Engl J Med. 2010; 362(11):986–993. [PubMed: 20237344]

Wand M, Marron J, Ruppert D. Transformation in density estimation (with comments). J Am Stat Assoc. 1991; 86:343–361.

Wang T, Gona P, Larson M, Tofler G, Levy D, Newton-Cheh C, Jacques P, Rifai N, Selhub J, Robins S. Multiple biomarkers for the prediction of first major cardiovascular events and death. N Engl J Med. 2006; 355:2631–2639. [PubMed: 17182988]

Wang R, Lagakos S, Ware J, Hunter D, Drazen J. Statistics in medicine-reporting of subgroup analyses in clinical trials. N Engl J Med. 2007; 357(21):2189–2194. [PubMed: 18032770]

Wilson PW, D'Agostino RB, Levy D, Belanger AM, Silbershatz H, Kannel WB. Prediction of cornary heart disease using risk factor categories. Circulation. 1998; 97:1837–1847. [PubMed: 9603539]

Zhao, L.; Cai, T.; Tian, L.; Uno, H.; Solomon, S.; Wei, L.; Minnier, J.; Kohane, I.; Pencina, M.; D'Agostino, R., et al. Harvard University Biostatistics Working Paper Series 2010: Working Paper 122. 2010. Stratifying subjects for treatment selection with censored event time data from a comparative study.

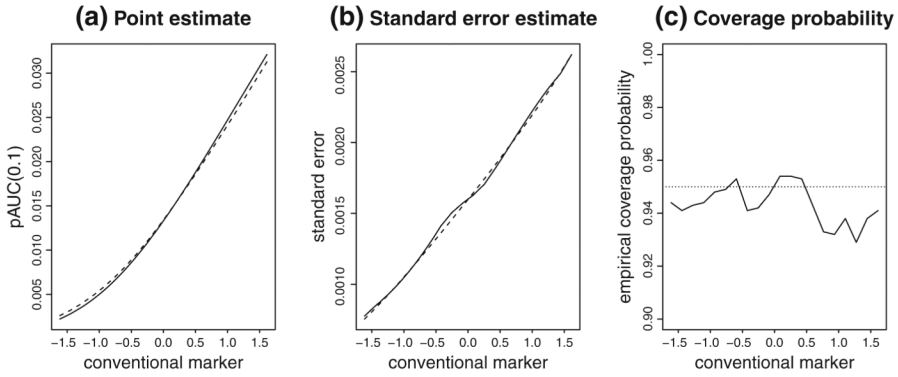

Fig. 1.

Performance of the point estimates, the standard error estimates and pointwise CIs for $pAUC_{0.1}$ with sample size 10,000: (**a**) the true $pAUC_{0.1}(x)$ (*solid*) and the average point estimates (*dashed*) over 1,000 replicates, (**b**) the empirical standard error estimates (*solid*) and the average of the estimated errors (*dashed*) based on the resampling procedure, and (**c**) the empirical coverage levels of the pointwise 95% CIs obtained from the resampling procedures

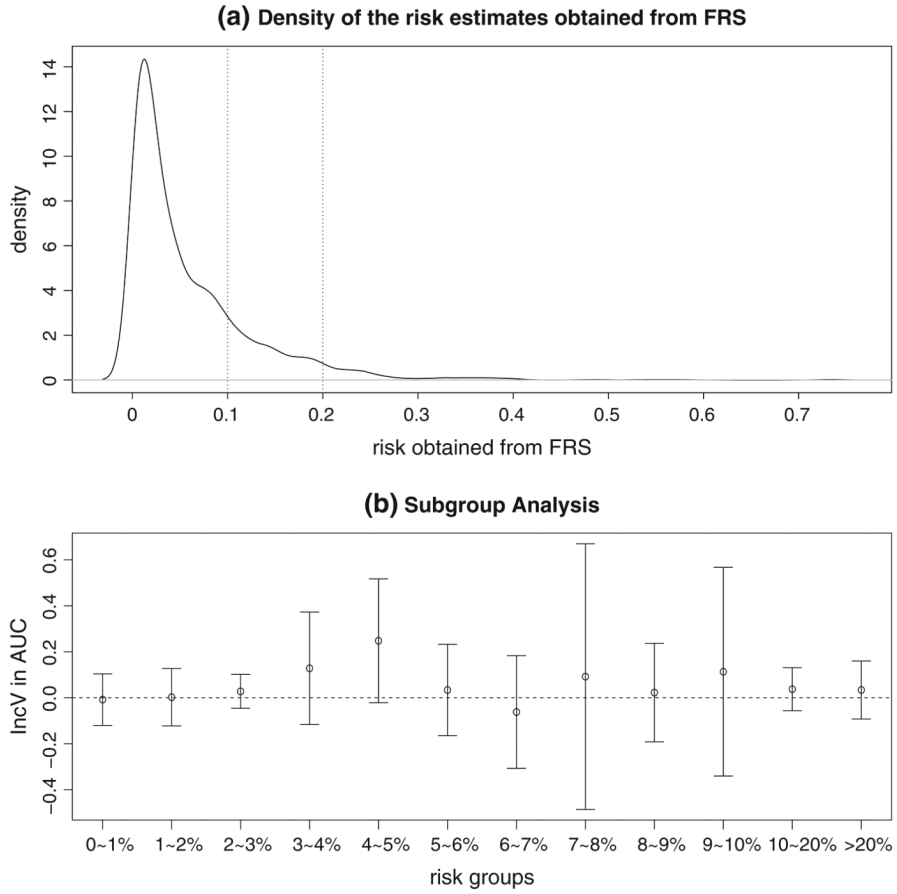

Fig. 2.

(**a**) The density estimates of the 10-year event risk calculated from the FRS. (**b**) The IncVs wrt C-statistics for the 10 subgroups of low risk as well as the intermediate and high risk groups with their 95% CIs

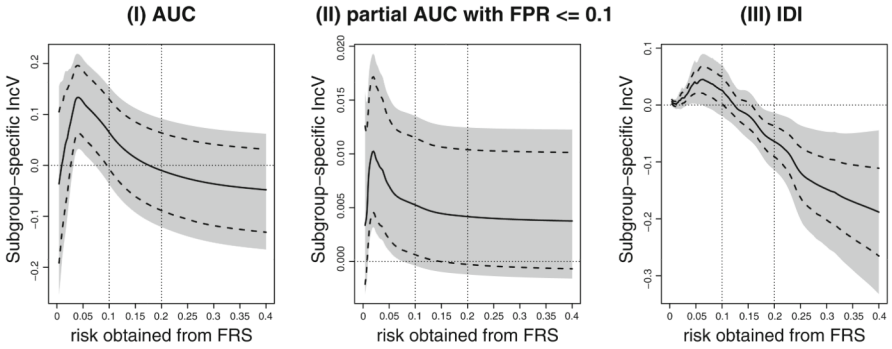

Fig. 3.

The point estimates (*solid line*), and its 95% pointwise CIs (*dashed lines*) and the 95% simultaneous confidence bands (*dark shaded region*) for (**I**) the subgroup-specific IncV with respect to AUC, $\mathrm{AUC}(x) - 1/2$; (**II**) the subgroup-specific IncV with respect to $\mathrm{pAUC}_{0.1}$, $\mathrm{pAUC}_{0.1}(x) - 0.1^2/2$; (**III**) the subgroup-specific IncV with respect to IDI. The two *vertical dotted lines* represent the-risk category cut-offs, 10 and 20%, from *left* to *right*

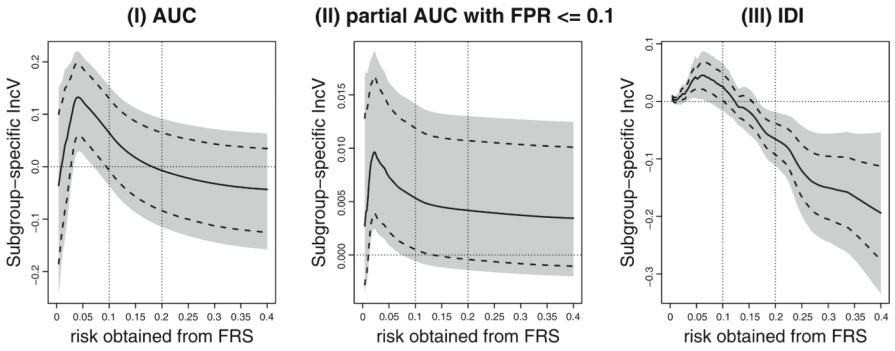

Fig. 4.

The point estimates (*solid line*), and its 95% pointwise CIs (*dashed lines*) and the 95% simultaneous confidence bands (*dark shaded region*) for the subgroup-specific IncV with respect to AUC and $pAUC_{0.1}$ as well as IDI. The results are based on the weights $\widehat{\omega}_i$ with $G_{X,Z}(t)$ estimated via a Cox PH model

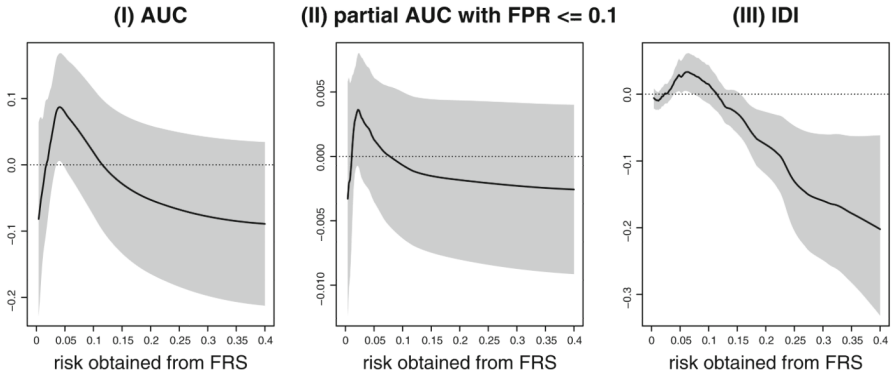

Fig. 5.

The point estimates (*solid line*), and its 95% simultaneous confidence bands (*dark shaded region*) for the relative subgroup-specific IncVs, which are used to test for heterogeneous IncVs, with respect to AUC and $pAUC_{0.1}$ as well as IDI over the interval [0,0.4] of the risk estimates obtained from the FRS