Harvard University
Harvard University Biostatistics Working Paper Series

Year 2010 *Paper* 123

Landmark Prediction of Survival

Layla Parast* Tianxi Cai†

*Harvard School of Public Health, lparast@hsph.harvard.edu
†Harvard School of Public Health, tcai@hsph.harvard.edu

Landmark Prediction of Survival

Layla Parast and Tianxi Cai
lparast@hsph.harvard.edu tcai@hsph.harvard.edu

Department of Biostatistics, Harvard School of Public Health, 677 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

Summary. Recent advancement in technology has lead to a wide range of genetic and biological markers that hold great potential in improving the prediction of survival outcomes. Although such new markers promise better disease prognosis, the accuracy in identifying short term and long term survivors remains unsatisfactory for most complex diseases. It has often been argued that short term clinical outcomes may have potential in predicting long term outcomes. In this paper, we propose to develop and evaluate conditional prognostic rules for the prediction of long term outcomes based on baseline marker information along with short term outcome status at an earlier landmark time. When there are multiple markers available, we construct an optimal composite score by fitting a proportional hazards working model for the conditional survival distribution. We also provide inference procedures for evaluating the incremental value of new markers in landmark prediction. The accuracy of the score is evaluated non-parametrically based on inverse probability weighting. Resampling procedures are proposed to derive estimation procedures for the accuracy measures. With a real example and numerical studies, we demonstrate that the proposed procedures perform well in finite samples.
Keywords: *biomarkers, disease prognosis, predictive accuracy, risk prediction, survival analysis.*

1. Introduction

In studies designed to develop prognostic classifiers based on predictive markers, marker measurements are often ascertained at baseline and patients are followed over time for the occurrence of certain clinical conditions or death. Since the risk for the disease occurrence may change over time, the time domain must be incorporated when developing prognostic rules. When there are multiple markers available to assist in prediction, it is of clinical interest to construct an optimal prognostic index based on available marker information. In the standard diagnostic setting with binary outcomes, various procedures have been proposed to combine multiple markers to improve diagnostic accuracy (Su and Liu, 1993; Pepe and Thompson, 2000; McIntosh and Pepe, 2002; Pepe et al., 2005). For event time outcomes, the most popular approach to combine markers for predicting time to disease onset is to fit the Cox proportional hazards model,

$$\lambda(t \mid \mathbf{Z}) = \lambda_0(t) \exp\{\boldsymbol{\beta}'\mathbf{Z}\},$$

and use $\widehat{\boldsymbol{\beta}}'\mathbf{Z}$ for prediction, where $\lambda(t \mid \mathbf{Z})$ is the hazard function for a subject with marker value $\mathbf{Z}$, $\lambda_0(\cdot)$ is some unknown baseline hazard function, and $\widehat{\boldsymbol{\beta}}$ is the maximum partial likelihood estimate of $\boldsymbol{\beta}$ (Kalbfleisch and Prentice, 2002). For example, the Framingham Risk Score (Wilson et al., 1998) for predicting cardiovascular failure was constructed based on such a method. Recently, Zheng et al. (2006) and Cai and Cheng (2008) showed that when the Cox model fails to hold, the risk score $\widehat{\boldsymbol{\beta}}'\mathbf{Z}$ may have poor accuracy in discriminating subjects with $T \leq t$ from those with $T > t$ for some t. To improve the prediction accuracy, Zheng et al. (2006) and Uno et al. (2007) proposed the use of time-varying coefficient models which allow for different composite scores for predicting short term and long term survival.

In general, patient populations may be heterogenous and represent a mixture of different subtypes of disease. As such, subjects with similar clinical symptoms may have drastically different disease outcomes. For example, breast cancer patients sharing the same clinical features, such as lymph node status and histological grade, may have significantly different survival profiles. Recent advancement in technology has lead to a wide range of genetic and biological markers that hold great potential in improving the prediction of survival outcomes. In a recent breast cancer gene expression study, van't Veer et al. (2002) and van de Vijver et al. (2002) developed a gene score for prognosis based on a 70-gene profile. The inclusion of the gene score was shown to have improved the prognostic accuracy over the traditional clinical markers in predicting survival (Uno et al., 2007). Although such new classifiers promise better disease prognosis, the accuracy in identifying short term vs. long term survivors remains unsatisfactory for most complex diseases. It has been often argued that short term clinical outcomes may have potential in predicting long term survival. For example, van der Sluis et al. (1997) studied the extent to which short-term outcomes can predict long-term outcomes for pediatric polytrauma patients. Weisner et al. (2003) demonstrated a clear association between short-term and long-term treatment success for alcohol addiction. To optimally select prevention and treatment strategy, it would be of great interest to develop comprehensive prognostic systems for patients that could make prediction about both the short term survival and the long term survival given the short term outcome. Such evaluations provide a more complete picture of the long term trajectory of disease progression and thus can be helpful for patients to make risk benefit decisions.

In this paper, we propose to develop conditional prognostic rules for the prediction of long term outcomes based on baseline marker information along with short term outcome. When the short term and long term outcomes are the same clinical event, Van Houwelingen

(2007) proposed robust estimation procedures for regression coefficients under a proportional hazards landmark prediction model. Here, we propose to extend the procedures given in Van Houwelingen (2007) for the setting where the short term outcome is based on an intermediate event that may be different from the event of interest for the long term outcome. For example, when selecting treatment options for breast cancer patients, it may be helpful to provide the chance of long term survival with and without conditioning on information regarding the short term outcome of distant metastasis-free survival. On the other hand, growing evidence suggests that complex gene and environment interactions underlie a number of diseases (Hunter, 2005). While developing such prediction models, it is crucial to realize that most simplistic statistical models are unlikely to capture the true relationship between the event times and the predictors of interest. We propose robust inference procedures for making inference about model parameters without requiring correct specification of the model. In addition, we propose non-parametric model free procedures to assess the prediction performance of the risk score obtained from the landmark models. Procedures for evaluating the incremental value of new markers for landmark prediction are also derived. Simulation studies suggest that the proposed inference procedures perform well in finite sample and prediction rules obtained based on the robust landmark model outperform those derived from a global Cox model when the underlying patient populations are heterogeneous. Our procedures are illustrated using a breast cancer gene expression study.

2. Evaluating Conditional Prognostic Rules

Let $T_{\mathbb{L}}$ denote event time for the long term outcome and $T_{\mathbb{S}}^*$ denote the event time for the short term outcome, which may or may not be the same as $T_{\mathbb{L}}$. For example, $T_{\mathbb{L}}$ may represent time to death and $T_{\mathbb{S}}^*$ may represent time to distant metastasis. Due to the potential difference in the underlying disease process, patients who have a good short term outcome may have very different clinical outcomes from the general patient population. It is thus of interest to incorporate information on the short term outcome into the prediction of long term outcomes. Here, we are particularly interested in the prediction of $T_{\mathbb{L}}$ among subjects with a good short term outcome, defined as $\Omega_{t_0} = \{T_{\mathbb{S}}^* > t_0, T_{\mathbb{L}} > t_0\} = \{T_{\mathbb{S}} > t_0\}$, where $T_{\mathbb{S}} = \min(T_{\mathbb{S}}^*, T_{\mathbb{L}})$. Such a rule could be used to distinguish subjects who will fail within τ years since t_0 from those who will survive $\tau + t_0$ years among Ω_{t_0}, where $\text{pr}(X_{\mathbb{L}} > \tau + t_0, X_{\mathbb{S}} > t_0) > 0$. Note that the classification of $T_{\mathbb{L}} > \tau + t_0$ for subjects in Ω_{t_0} is

equivalent to the classification of whether the residual life $R_{\mathbb{L}}^{t_0} = T_{\mathbb{L}} - t_0$ is greater than τ. Let $\mathbf{Z}$ denote the $p \times 1$ vector of predictors measured at baseline and let $\mathcal{Z}_\eta = \eta(\mathbf{Z})$ denote the composite score based on $\mathbf{Z}$ for predicting the long term outcome: $D_{\mathbb{L}}^{t_0+\tau} = I(T_{\mathbb{L}} \leq \tau + t_0) = I(R_{\mathbb{L}}^{t_0} \leq \tau)$.

For any given score $\mathcal{Z}_\eta$, one may evaluate its potential in predicting $D_{\mathbb{L}}^{\tau}$ by extending various time-specific accuracy measures as suggested in the literature (Heagerty et al., 2000; Cai et al., 2006; Uno et al., 2007; Cai and Cheng, 2008) to incorporate the conditional prognosis given $T_{\mathbb{S}} > t_0$. For example, the discrimination accuracy of $\mathcal{Z}_\eta$ for classifying $D_{\mathbb{L}}^{t_0+\tau}$ among Ω_{t_0} can be summarized by the time-specific sensitivity and specificity functions:

$$\text{Sens}_{t_0,\tau}(c) = \text{pr}_{\Omega_{t_0}}(\mathcal{Z}_\eta > c \mid D_{\mathbb{L}}^{t_0+\tau} = 1), \quad \text{and} \quad \text{Spec}_{t_0,\tau}(c) = \text{pr}_{\Omega_{t_0}}(\mathcal{Z}_\eta < c \mid D_{\mathbb{L}}^{t_0+\tau} = 0)$$

where $\text{pr}_{\Omega_{t_0}}$ represents the probability taken over the sub-population in Ω_{t_0}. To adequately summarize the inherent discrimination accuracy for a score, $\text{Sens}_{t_0,\tau}(\cdot)$ and $\text{Spec}_{t_0,\tau}(\cdot)$ must be considered simultaneously since higher values of $\text{Sens}_{t_0,\tau}(\cdot)$, obtained by lowering the threshold, are achieved at the expense of decreasing the $\text{Spec}_{t_0,\tau}(\cdot)$. A commonly used technique for summarizing the trade-offs between the sensitivity and specificity is the Receiver Operating Characteristic (ROC) curve (Swets and Pickett, 1982; Hanley, 1989; Begg, 1991). The time-specific ROC curve can be defined as a plot of: $\{1 - \text{Spec}_{t_0,\tau}(c), \text{Sens}_{t_0,\tau}(c), c \in (-\infty, \infty)\}$, or, equivalently, the function $\{u, \text{ROC}_{t_0,\tau}(u) = \text{Sens}_{t_0,\tau}(\text{Spec}_{t_0,\tau}^{-1}(1-u)), u \in (0,1)\}$. The overall accuracy of $\mathcal{Z}_\eta$ is often summarized by the area under the ROC curve (AUC), $\text{AUC}_{t_0,\tau} = \int \text{ROC}_{t_0,\tau}(u)du$.

These classification accuracy measures are useful for describing the inherent capacity a score has in discriminating $D_{\mathbb{L}}^{t}$ and thus for identifying optimal scoring systems and developing rules for assigning subjects into good or poor prognostic groups based on a selected threshold value c. After such a rule is identified, it would be important to examine the survival probability for patients assigned into the good or poor prognosis groups. Such probabilities are often summarized based on the positive predictive values (PPV) and negative predictive values (NPV), defined as

$$\text{PPV}_{t_0,\tau}(c) = \text{pr}_{\Omega_{t_0}}(D_{\mathbb{L}}^{t_0+\tau} = 1 \mid \mathcal{Z}_\eta > c), \quad \text{and} \quad \text{NPV}_{t_0,\tau}(c) = \text{pr}_{\Omega_{t_0}}(D_{\mathbb{L}}^{t_0+\tau} = 0 \mid \mathcal{Z}_\eta \leq c).$$

3. Developing Conditional Prognostic Rules

3.1. Models for constructing a composite score

When $T_{\mathbb{L}} = T_{\mathbb{S}}$, various standard survival models could be used for constructing composite scores to predict $D_{\mathbb{L}}^{t_0+\tau}$ among those with $T_{\mathbb{S}} > t_0$. One simple approach is to employ a

global proportional hazards (PH) model

$$\mathrm{pr}(T_{\mathbb{L}} > t_0 + \tau, T_{\mathbb{S}} > t_0 \mid \mathbf{Z}) = \mathrm{pr}(T_{\mathbb{L}} > t_0 + \tau \mid \mathbf{Z}) = \exp\{-\Lambda_0(t_0+\tau)\exp(\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Z})\} \quad (1)$$

where $\Lambda_0(\cdot)$ is an unspecified baseline cumulative hazard function and $\boldsymbol{\beta}_0$ is the unknown covariate effect. Under model (1), the conditional survival probability is

$$\mathrm{pr}_{\Omega_{t_0}}(T_{\mathbb{L}} > t_0 + \tau \mid \mathbf{Z}) = \mathrm{pr}_{\Omega_{t_0}}(R_{\mathbb{L}}^{t_0} > \tau \mid \mathbf{Z}) = \exp\{-\Lambda_{t_0}(\tau)\exp(\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Z})\}$$

where $\Lambda_{t_0}(\tau) = \Lambda_0(t_0+\tau) - \Lambda_0(t_0)$. This, together with the arguments given in McIntosh and Pepe (2002), implies that under the PH model, $\boldsymbol{\beta}_0^{\mathsf{T}}\mathbf{Z}$ is the optimal composite score for classifying $D_{\mathbb{L}}^{t_0+\tau}$ among Ω_{t_0}, for any given t_0 and τ. Here, the optimality is with respect to the ROC curve $\mathrm{ROC}_{t_0,\tau}(\cdot)$. This, together with the consistency of the maximum partial likelihood estimator of $\boldsymbol{\beta}_0$, $\widehat{\boldsymbol{\beta}}$, the estimated risk score $\widehat{\boldsymbol{\beta}}^{\mathsf{T}}\mathbf{Z}$ is the optimal score for predicting $D_{\mathbb{L}}^{t_0+\tau}$ asymptotically.

In general, when $T_{\mathbb{S}}$ and $T_{\mathbb{L}}$ represent survival times for two different outcomes, various bivariate survival models could be considered to construct an optimal score for the prediction of the residual life $R_{\mathbb{L}}^{t_0}$. Existing inference procedures for such conditional survival models are often derived based on joint inference on bivariate survival via bivariate copula or frailty modeling frameworks (Shih and Louis, 1995; Klaassen and Wellner, 1997; Oakes and Ritz, 2000; Pitt et al., 2006; Hougaard, 1995; Ha et al., 2001; Cai et al., 2002; Zeng and Lin, 2007). However, when the fitted model fails to hold, these procedures may not perform well yielding unstable prediction rules with poor predictive accuracy. To overcome such difficulties, we propose to construct time specific scores for any given t_0 and τ of interest. In particular, we propose to fit a conditional proportional hazards working model for the residual life $R_{\mathbb{L}}^{t_0}$ among Ω_{t_0}:

$$\mathrm{pr}_{\Omega_{t_0}}(R_{\mathbb{L}}^{t_0} > \tau \mid \mathbf{Z}) = \exp\left\{-\Lambda_0^{t_0}(\tau)\exp(\bar{\boldsymbol{\beta}}_{t_0}^{\mathsf{T}}\mathbf{Z})\right\} \quad (2)$$

where $\Lambda_0^{t_0}(\cdot)$ is the unspecified baseline cumulative hazard function among Ω_{t_0} and $\bar{\boldsymbol{\beta}}_{t_0}$ is the unknown covariate effect. When the model (2) holds, arguments as given above can be used to show that the binary classification rule $I(\mathcal{Z}_\eta > c)$ where $\mathcal{Z}_\eta = \exp\left\{-\Lambda_0^{t_0}(\tau)\exp(\bar{\boldsymbol{\beta}}_{t_0}^{\mathsf{T}}\mathbf{Z})\right\}$ has the optimal limiting ROC curve for classifying $D_{\mathbb{L}}^{t_0+\tau}$ among Ω_{t_0}. Note that model (2) includes the global Cox model (1) as a special case when $T_{\mathbb{L}} = T_{\mathbb{S}}$.

3.2. Inference Procedures for Model Parameters and Accuracy Measures

Due to censoring and competing risks, one may not observe $T_{\mathbb{L}}$ or $T_{\mathbb{S}}^*$ directly. In the presence of competing risks, the short term outcome $T_{\mathbb{S}}^*$ such as distant metastasis may not

be observable if $T_{\mathbb{L}}$ corresponds to a terminal event such as death. However, we assume that $T_{\mathbb{L}}$ is only subject to non-informative censoring, i.e. $T_{\mathbb{L}}$ will not be censored by $T_{\mathbb{S}}^*$. Under such assumptions, both $T_{\mathbb{L}}$ and $T_{\mathbb{S}} = \min(T_{\mathbb{S}}^*, T_{\mathbb{L}})$ are only subject to independent censoring. Thus for T_ι, we observe $X_\iota = \min(T_\iota, C)$, $\delta_\iota = I(T_\iota \leq C)$, for $\iota = \mathbb{S}, \mathbb{L}$, where C is time of censoring. We assume that C is independent of $(T_{\mathbb{L}}, T_{\mathbb{S}})$ and $\mathbf{Z}$ with a common survival function $G(t) = \text{pr}(C \geq t)$. Suppose data for analysis consist of n independent and identically distributed random vectors $\{(X_{\mathbb{L}i}, \delta_{\mathbb{L}i}, , X_{\mathbb{S}i}, \delta_{\mathbb{S}i}, \mathbf{Z}_i), i = 1, ..., n\}$.

To estimate $\bar{\boldsymbol{\beta}}_{t_0}$, we use the subgroup with $X_{\mathbb{S}} \geq t_0$ as an unbiased random sample for the subpopulation Ω_{t_0} since $\text{pr}(T_{\mathbb{L}} \leq t_0 + \tau, \mathbf{Z} \leq \mathbf{z} \mid X_{\mathbb{S}} \geq t_0) = \text{pr}_{\Omega_{t_0}}(R_{\mathbb{L}}^{t_0} \leq \tau, \mathbf{Z} \leq \mathbf{z})$. It follows that $\bar{\boldsymbol{\beta}}_{t_0}$ can be estimated by maximizing the partial likelihood function using subjects with $X_{\mathbb{S}} \geq t_0$. Specifically, we propose to obtain $\widehat{\boldsymbol{\beta}}_{t_0}$ as the maximizer of

$$\widehat{\ell}_{t_0}(\boldsymbol{\beta}) = \sum_{i:X_{\mathbb{S}i} \geq t_0} \int_{t_0}^{\infty} \left[\boldsymbol{\beta}^{\mathsf{T}} \mathbf{Z}_i - \log \left\{ \sum_{j:X_{\mathbb{S}j} > t_0} \exp\{\boldsymbol{\beta}^{\mathsf{T}} \mathbf{Z}_j\} I(X_{\mathbb{L}j} \geq s) \right\} \right] dN_i(s) \qquad (3)$$

where $N_i(s) = I(X_{\mathbb{L}i} \leq s)\delta_{\mathbb{L}i}$. In Appendix A, we show that $\widehat{\boldsymbol{\beta}}_{t_0} \to \boldsymbol{\beta}_{t_0}$ in probability as $n \to \infty$ regardless of the adequacy of (2), where $\boldsymbol{\beta}_{t_0}$ is the unique maximizer of the limiting objective function, $\ell_{t_0}(\boldsymbol{\beta})$. Furthermore, we show that $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{t_0} - \boldsymbol{\beta}_{t_0})$ converges in distribution to a multivariate normal with mean zero.

To evaluate our prediction rule, we propose the following estimators for the aforementioned accuracy measures using inverse probability weighting. Specifically, let

$$\widehat{\text{Sens}}_{t_0,\tau}(c) = \frac{\sum_{i:X_{\mathbb{S}i} > t_0} \widehat{W}_i I(\widehat{\boldsymbol{\beta}}_{t_0}^{\mathsf{T}} \mathbf{Z}_i \geq c) I(X_{\mathbb{L}i} \leq t_0 + \tau)}{\sum_{i:X_{\mathbb{S}i} > t_0} \widehat{W}_i I(X_{\mathbb{L}i} \leq t_0 + \tau)}$$

$$\widehat{\text{Spec}}_{t_0,\tau}(c) = \frac{\sum_{i:X_{\mathbb{S}i} > t_0} \widehat{W}_i I(\widehat{\boldsymbol{\beta}}_{t_0}^{\mathsf{T}} \mathbf{Z}_i < c) I(X_{\mathbb{L}i} > t_0 + \tau)}{\sum_{i:X_{\mathbb{S}i} > t_0} \widehat{W}_i I(X_{\mathbb{L}i} > t_0 + \tau)}$$

$$\widehat{\text{PPV}}_{t_0,\tau}(c) = \frac{\sum_{i:X_{\mathbb{S}i} > t_0} \widehat{W}_i I(\widehat{\boldsymbol{\beta}}_{t_0}^{\mathsf{T}} \mathbf{Z}_i \geq c) I(X_{\mathbb{L}i} \leq t_0 + \tau)}{\sum_{i:X_{\mathbb{S}i} > t_0} \widehat{W}_i I(\widehat{\boldsymbol{\beta}}_{t_0}^{\mathsf{T}} \mathbf{Z}_i \geq c)}$$

$$\widehat{\text{NPV}}_{t_0,\tau}(c) = \frac{\sum_{i:X_{\mathbb{S}i} > t_0} \widehat{W}_i I(\widehat{\boldsymbol{\beta}}_{t_0}^{\mathsf{T}} \mathbf{Z}_i < c) I(X_{\mathbb{L}i} > t_0 + \tau)}{\sum_{i:X_{\mathbb{S}i} > t_0} \widehat{W}_i I(\widehat{\boldsymbol{\beta}}_{t_0}^{\mathsf{T}} \mathbf{Z}_i < c)}.$$

where $\widehat{W}_i = \frac{I(X_{\mathbb{L}i} > t_0 + \tau)}{\widehat{G}(t_0 + \tau)} + \frac{I(X_{\mathbb{L}i} \leq t_0 + \tau)\delta_{\mathbb{L}i}}{\widehat{G}(X_{\mathbb{L}i})}$ and $\widehat{G}(\cdot)$ is the Kaplan-Meier estimator of $G(\cdot)$. Subsequently, $\text{ROC}_{t_0,\tau}(u)$ can be estimated as $\widehat{\text{ROC}}_{t_0,\tau}(u) = \widehat{\text{Sens}}_{t_0,\tau}\{\widehat{\text{Spec}}_{t_0,\tau}^{-1}(1-u)$. Similarly, $\text{AUC}_{t_0,\tau}$ can be estimated by

$$\widehat{\text{AUC}}_{t_0,\tau} = \frac{\sum_{i:X_{\mathbb{S}i} > t_0, X_{\mathbb{L}i} \leq t_0 + \tau} \sum_{j:X_{\mathbb{S}j} > t_0, X_{\mathbb{L}j} > t_0 + \tau} I(\widehat{\boldsymbol{\beta}}_{t_0}^{\mathsf{T}} \mathbf{Z}_i > \widehat{\boldsymbol{\beta}}_{t_0}^{\mathsf{T}} \mathbf{Z}_j) \widehat{W}_i \widehat{W}_j}{\sum_{i:X_{\mathbb{S}i} > t_0, X_{\mathbb{L}i} \leq t_0 + \tau} \sum_{j:X_{\mathbb{S}j} > t_0, X_{\mathbb{L}j} > t_0 + \tau} \widehat{W}_i \widehat{W}_j}$$

In Appendix B we show that the accuracy functions are consistent for the theoretical counterparts, uniformly in (c, t_0, τ). Furthermore, we show that $\sqrt{n}(\widehat{\text{Sens}}_{t_0,\tau}(c) - \text{Sens}_{t_0,\tau}(c))$ converges weakly to a zero-mean Gaussian processes. Similarly, this holds for the other accuracy measures and it is not difficult to show that these convergences hold jointly. It then follows from the stochastic equi-continuity of these processes that $\sqrt{n}\{\widehat{\text{ROC}}_{t_0,\tau}(u) - \text{ROC}_{t_0,\tau}(u)\}$ also converges weakly to a Gaussian process. Similar weak convergence results for other measures with cut-off values selected to achieve a sensitivity level of u, e.g. $\widehat{\text{PPV}}_{t_0,\tau}\{\text{Sens}^{-1}_{t_0,\tau}(u)\}$.

However, it is difficult to empirically estimate the variance associated with these processes as they involve derivative functions that are difficult to estimate especially under model mis-specification. To overcome such difficulties, we propose using a perturbation-resampling method (Park and Wei, 2003; Cai et al., 2005; Tian et al., 2007) to approximate the distributions of the proposed estimators. In particular, let $\{V_i^{(b)}, i = 1, ..., n, b = 1, ..., B\}$ be nB independent realizations of a positive random variable V from a known distribution with unit mean and unit variance. Let $\widehat{\boldsymbol{\beta}}^{(b)}_{t_0}$ be the minimizer of $\widehat{\ell}^{(b)}_{t_0}(\boldsymbol{\beta}) = \sum_{i:X_{\mathbb{S}i}>t_0}[\int\{\boldsymbol{\beta}^{\intercal}\mathbf{Z}_i - \log\sum_{j:X_{\mathbb{S}j}>t_0} V_j^{(b)}\exp\{\boldsymbol{\beta}^{\intercal}\mathbf{Z}_j\}I(X_{\mathbb{L}j}\geq s)\}V_i^{(b)}dN_i(s)]$. For accuracy measures such as $\text{AUC}_{t_0,\tau}$ and $\text{Sens}_{t_0,\tau}(c)$, let

$$\widehat{\text{AUC}}^{(b)}_{t_0,\tau} = \frac{\sum_{i:X_{\mathbb{S}i}>t_0, X_{\mathbb{L}i}\leq t_0+\tau}\sum_{j:X_{\mathbb{S}j}>t_0, X_{\mathbb{L}j}>t_0+\tau} I(\mathbf{Z}_i^{\intercal}\widehat{\boldsymbol{\beta}}^{(b)} > \mathbf{Z}_j^{\intercal}\widehat{\boldsymbol{\beta}}^{(b)})\widehat{W}_i^{(b)}\widehat{W}_j^{(b)}V_i^{(b)}V_j^{(b)}}{\sum_{i:X_{\mathbb{S}i}>t_0, X_{\mathbb{L}i}\leq t_0+\tau}\sum_{j:X_{\mathbb{S}j}>t_0, X_{\mathbb{L}j}>t_0+\tau}\widehat{W}_i^{(b)}\widehat{W}_j^{(b)}V_i^{(b)}V_j^{(b)}}$$

$$\widehat{\text{Sens}}^{(b)}_{t_0,\tau}(c) = \frac{\sum_{i:X_{\mathbb{S}i}>t_0, X_{\mathbb{L}i}\leq t_0+\tau}\widehat{W}_i^{(b)}I(\mathbf{Z}_i^{\intercal}\widehat{\boldsymbol{\beta}}^{(b)}_{t_0}\geq c)V_i^{(b)}}{\sum_{i:X_{\mathbb{S}i}>t_0, X_{\mathbb{L}i}\leq t_0+\tau}\widehat{W}_i^{(b)}V_i^{(b)}}$$

where $\widehat{W}_i^{(b)} = I(X_{\mathbb{L}i} > t_0+\tau)/\widehat{G}^{(b)}(t_0+\tau) + I(X_{\mathbb{L}i}\leq t_0+\tau)\delta_{\mathbb{L}i}/\widehat{G}^{(b)}(X_{\mathbb{L}i})$ and $\widehat{G}^{(b)}$ is the Kaplan-Meier estimator of $G(\cdot)$ with weights $V_i^{(b)}$. The empirical distributions of these realizations can be used to approximate the distribution of the corresponding estimators. For example, the variance of $\widehat{\text{AUC}}_{t_0,\tau}$ can be estimated using

$$\widehat{\sigma}^2_{t_0,\tau} = B^{-1}\sum_{b=1}^{B}\{\widehat{\text{AUC}}^{(b)}_{t_0,\tau} - \widehat{\text{AUC}}_{t_0,\tau}\}^2.$$

A $100(1-2\alpha)\%$ confidence interval (CI) for $\text{AUC}_{t_0,\tau}$ could be obtained either using the normal CI or the empirical quantiles of the perturbed samples. The validity of the resampling procedure can be justified based on the large sample theory given in Appendix B along with similar arguments as given in Cai et al. (2005).

The aforementioned accuracy estimators, commonly referred to as apparent accuracy estimates, tend to be overly optimistic, particularly when the dimension of $\mathbf{Z}$ is not small.

To reduce this potential bias, we consider a general cross-validation where we randomly split the data into a training set of size n_t and a validation set of size $n_v = n - n_t$. For the kth split, we estimate $\boldsymbol{\beta}_{t_0}$ using the training set, denoted by $\widehat{\boldsymbol{\beta}}_{t_0}^{(\text{CV})_k}$. Then for any accuracy measure of interest, say $\text{AUC}_{t_0,\tau}$, we estimate the AUC for the linear score $\widehat{\boldsymbol{\beta}}_{t_0}^{(\text{CV})_k\text{T}}\mathbf{Z}$ based on the validation set, denoted by $\widehat{\text{AUC}}_{t_0,\tau}^{(\text{CV})_k}$. Let $\widehat{\text{AUC}}_{t_0,\tau}^{(\text{CV})}$ be the average of all $\widehat{\text{AUC}}_{t_0,\tau}^{(\text{CV})_k}$ over K random splits. Using similar justification as given by Tian et al. (2007) and Uno et al. (2007), it can be shown that $\widehat{\text{AUC}}_{t_0,\tau}^{(\text{CV})}$ and $\widehat{\text{AUC}}_{t_0,\tau}$ have the same limiting distribution at the first order. Therefore, one may construct a 95% CI for $\text{AUC}_{t_0,\tau}$ as $\widehat{\text{AUC}}_{t_0,\tau}^{(\text{CV})} \pm 1.96\widehat{\sigma}_{t_0,\tau}$.

It is important to note that our proposed method holds for any choice of t_0 and τ. In practice, one may consider the proposed estimator across a range of t_0 and τ as a basis for selecting an appropriate time point such that the prediction model is most accurate. For example, suppose t_0 is pre-selected by a clinician, such as a regularly scheduled 1-year appointment. Information guiding the choice of τ would be a clinically meaningful as it would shed light on what time point/period $\mathbf{Z}$ is most useful in prediction. To determine the values of τ that offer high overall accuracy from classifying $D_{\mathbb{L}}^{t_0+\tau}$ for a fixed t_0, one may plot $\widehat{\text{AUC}}_{t_0,\tau_j}^{(\text{CV})}$ over a range of τ_j in some interval $[\tau_l, \tau_r] \subset (t_0, \max(X_{\mathbb{L}i}))$ and construct both point-wise and simultaneous CIs using the perturbation-resampling method as described above. A $100(1-\alpha)\%$ simultaneous CI for $\{\widehat{\text{AUC}}_{t_0,\tau_j}^{(\text{CV})}, \tau_j \in [\tau_l, \tau_r]\}$ can be obtained as $\{\widehat{\text{AUC}}_{t_0,\tau_j}^{(\text{CV})} \pm \widehat{c}_\alpha \widehat{\sigma}_{t_0,\tau}, \tau_j \in [\tau_l, \tau_r]\}$, where $\widehat{c}_\alpha$ is the $(1-\alpha)$ empirical quantile of $\{\sup_{\tau_j \in [\tau_l,\tau_r]} |\widehat{\text{AUC}}_{t_0,\tau_j}^{(b)} - \widehat{\text{AUC}}_{t_0,\tau_j}| / \widehat{\sigma}_{t_0,\tau_j}, b = 1, ..., B\}$. The simultaneous CI will ensure the control of family-wise type I error when selecting a set of τ_j's such that the overall accuracy is above a certain threshold value.

3.3. Incremental Value of New Markers

Determining the incremental value (IV) of new markers in prediction is often of clinical interest, particularly if measuring the markers is expensive or invasive. Several procedures have been developed to quantify the overall IV from a new marker for the entire population (Tian et al., 2007; Uno et al., 2007; Pepe et al., 2004, 2008; Pencina et al., 2008). Our proposed method could potentially shed light on how the IV may vary across sub-populations. For example, if the new marker is only useful for predicting $T_{\mathbb{L}}$ among those with good prognosis, one may expect that the IV is near zero for predicting $T_{\mathbb{S}} \geq t_0$, but is high for predicting $T_{\mathbb{L}} < t_0 + \tau$ among those with $T_{\mathbb{S}} > t_0$. This may provide a useful tool for practitioners to decide when the new marker is needed in addition to conventional risk factors. Let (M_1) denote the model with routine markers only and (M_2) denote the model with both

routine markers and the new markers. One may quantify the IV of the new markers based on $\Delta_{t_0,\tau} = \mathrm{AUC}^{(M_2)}_{t_0,\tau} - \mathrm{AUC}^{(M_1)}_{t_0,\tau}$, where $\mathrm{AUC}^{(M_k)}_{t_0,\tau}$ denotes the AUC of (M_k), for $k = 1, 2$. To make inference about $\Delta_{t_0,\tau}$, let $\widehat{\mathrm{AUC}}^{(\mathrm{CV})(M_k)}_{t_0,\tau}$ denote the cross-validated estimate of $\mathrm{AUC}^{(M_k)}_{t_0,\tau}$ and $\{\widehat{\mathrm{AUC}}^{(M_k)(\mathrm{b})}_{t_0,\tau}, b = 1, ..., B\}$ be the perturbed estimates of $\mathrm{AUC}^{(M_k)}_{t_0,\tau}$. Then one may make inference about $\Delta_{t_0,\tau}$ based on $\hat{\Delta}^{(\mathrm{CV})}_{t_0,\tau} = \widehat{\mathrm{AUC}}^{(\mathrm{CV})(M_2)}_{t_0,\tau} - \widehat{\mathrm{AUC}}^{(\mathrm{CV})(M_1)}_{t_0,\tau}$ and the empirical variance of $\{B^{-1}\sum_{i=1}^{n}\{\widehat{\mathrm{AUC}}^{(M_2)(\mathrm{b})}_{t_0,\tau} - \widehat{\mathrm{AUC}}^{(M_1)(\mathrm{b})}_{t_0,\tau}, b = 1, ..., B\}$. To examine how the IV vary over τ, one may assess $\Delta_{t_0,\tau}$ as a function of τ and identify the range of τ such that M_2 is better than M_1. To this end, one may construct a plot of $\widehat{\Delta}^{(\mathrm{CV})}_{t_0,\tau_j} = \widehat{\mathrm{AUC}}^{(\mathrm{CV})(M_2)}_{t_0,\tau_j} - \widehat{\mathrm{AUC}}^{(\mathrm{CV})(M_1)}_{t_0,\tau_j}$. Simultaneous confidence intervals for $\{\Delta_{t_0,\tau_j}, \tau_j \in [\tau_l, \tau_r]\}$ could be constructed using similar procedure as for $\{\widehat{\mathrm{AUC}}^{(\mathrm{CV})}_{t_0,\tau_j}, \tau_j \in [\tau_l, \tau_r]\}$.

3.4. Comprehensive Prognosis of Short Term and Long Term Outcomes

In practice, one may be interested in both the prognosis of the short term and the long term outcomes. Our proposed procedures can easily be extended to make such joint predictions. Let $D^{t_0}_{\mathbb{S}} = I(T_{\mathbb{S}} \le t_0)$ denote the status of the short term outcome. When the underlying patient population consists of a mixture of short term survivors and long term survivors, the optimal prediction score for short term outcomes may be different from that of long term outcomes. Thus, to construct a robust prediction rule for $D^{t_0}_{\mathbb{S}}$, one may fit a separate Cox model using survival information on $T_{\mathbb{S}}$ up to t_0 as in Cai et al. (2010). Let $\widehat{\alpha}_{t_0}$ denote the maximum partial likelihood estimator for the effect of $\mathbf{Z}$ on $T_{\mathbb{S}}$ based on the truncated data $\{(X_{\mathbb{S}i} \wedge t_0, \delta_{\mathbb{S}i} I(X_{\mathbb{S}i} \le t_0), \mathbf{Z}_i), i = 1, ..., n\}$. Such a fitting essentially corresponds to assuming the Cox model holds up to t_0. For future subjects with outcomes $(T^0_{\mathbb{S}}, T^0_{\mathbb{L}})$ and covariate level $\mathbf{Z}^0$, we will classify them as having a poor short term prognosis, denoted by $\mathbb{S}^+_{t_0}$, if $\widehat{\alpha}^{\top}_{t_0}\mathbf{Z}^0 > \widehat{c}_{\mathbb{S}_{t_0}}$, and good short term prognosis, denoted by $\mathbb{S}^-_{t_0}$, otherwise. Similarly, for the long term conditional survival, we classify subjects as having a poor prognosis, denoted by $\mathbb{L}^{\tau+}_{t_0}$, if $\widehat{\boldsymbol{\beta}}^{\top}_{t_0}\mathbf{Z}^0 > \widehat{c}_{\mathbb{L}^{\tau}_{t_0}}$, and good prognosis, denoted $\mathbb{L}^{\tau-}_{t_0}$, otherwise. Here $\widehat{c}_{\mathbb{S}_{t_0}}$ and $\widehat{c}_{\mathbb{L}^{\tau}_{t_0}}$ are the cut-off values selected to achieve certain desired sensitivity or specificity levels for the classification of the corresponding events.

To illustrate how the proposed procedures may be useful in clinical practice, we next describe how a clinician may provide a two-step prognosis for future subjects. For future patients with $\mathbf{Z}^0$, the clinician will first classify them as either $\mathbb{S}^+_{t_0}$ or $\mathbb{S}^-_{t_0}$ and provide their chance of surviving to t_0 given their prognosis based on $\mathrm{pr}(T^0_{\mathbb{S}} > t_0 \mid \mathbb{S}^-_{t_0})$ and $\mathrm{pr}(T^0_{\mathbb{S}} > t_0 \mid \mathbb{S}^+_{t_0})$. In the next step, the clinician will further classify subjects as having a good or poor conditional prognosis and inform them regarding their chance of surviving an additional τ

years after t_0 provided that they do survive t_0 years i.e. $T_{\mathbb{S}}^0 > t_0$. This information would be based on $\text{pr}_{\Omega_{t_0}}(T_{\mathbb{L}}^0 > \tau + t_0 | \mathbb{S}_{t_0}^{\jmath}, \mathbb{L}_{t_0}^{\tau\jmath})$, for $\jmath = +$ and $-$. These conditional probabilities can be estimated non-parametrically via IPW similar to those given in section 3.2

4. Simulations

To examine finite-sample properties of the proposed estimation procedures, we conducted simulation studies under various scenarios and focused on the case with $T_{\mathbb{L}} = T_{\mathbb{S}} = T$ for simplicity. Two types of models were used to generate T: a mixture of log-normal in setting (i) and a Cox model in setting (ii). Two types of censoring patterns were considered in setting (i) to examine the effect of censoring. The covariate vector $\mathbf{Z}$ consists 5 components: $(Z_2, Z_3, Z_4) \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ with unit variance and a weak correlation ranging from -.03 and .05; $Z_1 \sim$ a bernoulli distribution with success probability $\Phi(Z_2 + Z_3)$; and $Z_5 \sim \text{Uniform}[\Phi(Z_4), 1 + \Phi(Z_4)]$. In each setting, we generated 1000 realizations of size n for n=200, 500 and 1,000 and obtained $\widehat{\boldsymbol{\beta}}_{t_0}$ as well as $\widehat{\boldsymbol{\beta}}$, the standard maximum partial likelihood estimator of $\boldsymbol{\beta}$ under a global Cox model. Accuracy measures were also calculated for both $\boldsymbol{\beta}$ estimates. Here, in addition to assessing the overall accuracy based on the AUC, we consider the specificity, PPV and NPV at a cut-off value corresponding to a sensitivity level of 0.90. The standard error estimates were obtained based on 500 perturbations each.

In setting (i), the survival time was generated from $T = \exp\{(\boldsymbol{\beta}_1^{\intercal}\mathbf{Z} + \epsilon_1)/6\}$ with $\epsilon_1 \sim N(-3, 1)$ if $\mathcal{B} = 1$ and $T = \exp\{(\boldsymbol{\beta}_0^{\intercal}\mathbf{Z} + \epsilon_0)/6\} + 1$ with $\epsilon_0 \sim N(10, 4)$ if $\mathcal{B} = 0$, where $\mathcal{B} \sim$ Bernoulli(.4), $\boldsymbol{\beta}_1 = (3, 6, 1.5, 0, 0)^{\intercal}$ and $\boldsymbol{\beta}_0 = c(0, 0, 1.5, 3, 1.5)$. This mimics a clinical setting in which (Z_1, Z_2, Z_3) are predictive of short term survival, (Z_3, Z_4, Z_5) are predictive of long term survival. For illustration, we chose t_0 to be year 1 to reflect an early indication of disease and τ to be year 5. Therefore, among patients surviving 1 year, we wish to estimate their probability of survival past year 6, given baseline covariates. Under these conditions, $\text{pr}(T \leq t_0) = .23$ and $\text{pr}(T \leq \tau + t_0 \mid T > t_0) = .49$. We first consider the case when C was generated from exponential with rate .11, yielding a censoring rate of approximately 40%. The results are shown in Tables **??**. Since both the global and the conditional Cox model are mis-specified in this case, $\widehat{\boldsymbol{\beta}}_{t_0}$ and $\widehat{\boldsymbol{\beta}}$ are converging to two different limits. $\widehat{\boldsymbol{\beta}}_{t_0}$ assigns more weights on (Z_3, Z_4, Z_5), the covariates that are helpful for predicting long term survival. On the contrary, $\widehat{\boldsymbol{\beta}}$ assigns more similar weights to all covariates. The resulting linear score has a higher overall accuracy with $\text{AUC}_{t_0,\tau} \sim .74$ for the landmark method compared to $\text{AUC}_{t_0,\tau} \sim .67$ using the global Cox model. All point estimators have negligible bias. The

standard errors estimated from the resampling method are close to the empirical standard errors and the empirical coverage levels are close to their nominal level.

We also considered a different setting to asses how censoring patterns effect the estimation. In this setting, T was generated from the same log-normal mixture distribution as described above. However, the censoring time was generated from a mixture of exponential(rate=.30) with probability .40 and exponential(rate =.02) with probability .60. Under this setting, about 10% are censored within the first year and 25% are censored within the first 6 years. As shown in Tables 2, $\widehat{\boldsymbol{\beta}}_{t_0}$ and $\widehat{\boldsymbol{\beta}}$ have limits different from those for the previous setting with exponential censoring. This is due to the fact that the maximizer of the partial likelihood function is no longer free of the censoring distribution under model mis-specification. The accuracy of the linear scores is also slightly different for these two settings. However, the proposed procedures remain to perform well with negligible bias in the point estimates and proper coverage level for the interval estimates.

In setting (ii), we are interested in examining the efficiency loss due to the use of the more robust conditional landmark model when the global Cox model holds. To this end, we generated T from a Cox model with $1.25\log(T) = 2 + \boldsymbol{\beta}_0^{\intercal}\mathbf{Z} + \epsilon$, where ϵ is from an extreme value distribution and $\boldsymbol{\beta}_0 = (1, .5, .5, 1, .5)^{\intercal}$. Under this configuration, $\text{pr}(T \leq t_0) = .20$, $\text{pr}(t_0 < T < \tau + t_0) = .33$ and $\text{pr}(T < \tau + t_0 \mid T > t_0) = .34$. Censoring time was generated from an exponential distribution with rate .05, yielding approximately 40% of censoring. Results shown in Tables 3. Since $\widehat{\boldsymbol{\beta}}$ is semi-parametric efficient and $\widehat{\boldsymbol{\beta}}_{t_0}$ is obtained based on the subset in Ω_{t_0}, which consists of 83% of the entire sample, there is a significant efficiency loss due to landmarking. However, the efficiency loss in estimating the accuracy measures is negligible.

5. Example

In this section, we illustrate our proposed procedures using a dataset originally used in van de Vijver et al. (2002) to evaluate a 70-gene risk score for breast cancer prognosis. More recently, Carter et al. (2006) demonstrated that a chromosome instability genetic score, denoted by CIN25, is predictive of survival for various types of cancer. Here we investigate the predictive ability of CIN25 for breast cancer survival using the data from van de Vijver et al. (2002). This dataset consists of 260 women, with individual information on time to death, time to distant metastasis, CIN25 gene score, age, tumor grade, size of tumor, baseline lymph node status, and estrogen receptor status.

For illustration, we first consider the case with $T_{\mathbb{L}} = T_{\mathbb{S}}^*$ being metastasis free survival. There were a total of 88 events and the 5-year event rate is about 0.24. To provide a comprehensive prediction for these patients, we first aim to make a prediction for a short term outcome, progression-free survival by t_0=3, i.e. $D_{\mathbb{S}}^{t_0} = I(T_{\mathbb{S}} \leq 3)$. The 3-year event rate was about 0.17. To construct a robust score for this prediction, we fit a Cox model using survival information up to t_0 and obtain a linear score based on such a truncated model. For comparison, we also obtain the linear score from fitting a global Cox model. After the prediction rule for $D_{\mathbb{S}}^{t_0}$ is developed and evaluated, we next construct and evaluate prediction rules for 5-year progression-free survival among among those who survived 3 years without metastasis based on our proposed landmark procedures. In Table 4(a), we present the regression coefficients estimates for the risk scores. For the short term prediction, the global Cox model assigns the highest weight on the gene score whereas the truncated model assigns more weight on the tumor grade. For the landmark prediction, the conditional Cox model also assigns the most weight on the gene score. This suggests that the gene score may be more useful for long term prediction than for short term prediction. As shown in Table 4(b), the risk score has a reasonable accuracy in classifying $T_{\mathbb{L}} \leq 3$ with AUC 0.733 (s.e. 0.035) for the truncated model and 0.728 for the global Cox model. To develop a prognostic rule for the short term survival, we select the cut-off value to achieve a sensitivity level of 0.90 and classify patients as having good prognosis if the predicted risk of failure is lower than the cut-off value. The rule from the truncated model leads to 24% chance of failure among those classified as poor prognosis and 96% of survival among those with good prognosis. Now among those who do survive 3-years without metastasis, the risk score has lower accuracy in predicting the long term outcome with AUC 0.64 for the landmark model and 0.63 for the global Cox model. At sensitivity level of 0.90, the landmark model yields a rule with specificity of 0.35 (s.e. 0.11). Based on the corresponding prognostic rule, the chance of survival is 97% for those with good long term prognosis and 77% for those with poor prognosis.

Now, to evaluate the incremental value of the gene score, we compared the aforementioned accuracy to those obtained by fitting the models with predictors without the gene score. For the short term outcome based on the truncated Cox model, the cross-validated estimates of AUC are 0.73 and 0.74 with and without the gene score, respectively. This again suggests that the gene score may not be useful for predicting short term survival. For the landmark prediction based on the conditional Cox model, the AUC estimate decreases from 0.64 to 0.52 when the gene score is removed from the model. A 95% confidence interval

for difference in the AUC is (0.02,0.22) suggesting that the gene score significantly improves the accuracy for the landmark prediction. To evaluate the IV of gene score for various values of τ, let M_2 and M_1 indicate models with and without the gene score, respectively. As shown in Figure 2, the IV of gene score slightly increases for larger values of τ and is significant for a few τ_j's. However, after adjusting for overall type I error based on the simultaneous band, the improvement due to the gene is no longer statistically significant.

To develop a comprehensive prognosis system, we then classify subjects as having a good or poor prognosis for both the short term outcome and the conditional long term outcome using all predictors. The cut-off values $\widehat{c}_{\mathbb{L}^\tau_{t_0}}$ and $\widehat{c}_{\mathbb{S}_{t_0}}$ are selected to achieve 90% sensitivity in classifying $D_{\mathbb{L}}^{t_0+\tau}$ among Ω_{t_0} and $D_{\mathbb{S}}^{t_0}$, respectively. As shown in Table 4(c), the chance of surviving 3 years without distant metastasis is 76% for those classified as $\mathbb{S}^+_{t_0}$, and 96% for those classified as $\mathbb{S}^-_{t_0}$. Now, if a patient does survive 3 years without metastasis, then her chance of surviving another 5 years is 97.7% if she is classified as $(\mathbb{S}^-_{t_0}, \mathbb{L}^{\tau-}_{t_0})$, 83.7% if she is classified as $(\mathbb{S}^-_{t_0}, \mathbb{L}^{\tau+}_{t_0})$, 95.8% if she is classified as $(\mathbb{S}^+_{t_0}, \mathbb{L}^{\tau-}_{t_0})$, 73.6% if she is classified as $(\mathbb{S}^+_{t_0}, \mathbb{L}^{\tau+}_{t_0})$.

We also consider the case with $T^*_{\mathbb{S}}$ corresponding to distant metastasis and $T_{\mathbb{L}}$ being overall survival and the goal is now prediction of 5-year overall survival among subjects with $T_{\mathbb{S}} = \min(T^*_{\mathbb{S}}, T_{\mathbb{L}}) > 3$ years. The results are summarized in Table 5. The prediction with both clinical and gene score has an AUC of 0.68 (se 0.048) which is slightly higher than that for the prediction of metastasis-free survival. The global Cox method yielded similar accuracy. The inclusion of gene score resulted in an increase of 0.09 (se 0.07) which is similar to that of metastasis-free survival.

6. Remarks

In this article we propose robust procedures for developing and evaluating conditional prognostic rules for the prediction of long term outcomes based on baseline marker and short term outcome information in order to improve prediction accuracy for long-term survivors. The proposed procedures yield stable prediction rules regardless of model adequacy. Such a robustness property is particularly important when $T_{\mathbb{L}}$ and $T_{\mathbb{S}}$ represent two different outcomes as for such settings, it is difficult if not impossible to identify bivariate survival models that capture the complex relationship between the correlated outcomes and the predictors. Under model mis-specification, traditional procedures for making inference may not be valid and thus lead to prediction rules that are either unstable or have unsatisfac-

tory accuracy. Furthermore, our proposed non-parametric procedures for making inference about the accuracy measures are valid regardless of model adequacy.

It is important to emphasize that we do not require the correct specification of the working model. In addition, the proposed procedure provides a prediction rule at baseline using only covariate information, $\mathbf{Z}$, available at baseline. If time-varying covariates are available, it would be of interest to provide an updated prediction rule at the landmark time t_0 using covariate information collected up to t_0. Our proposal can be extended by replacing $\boldsymbol{\beta}^\top \mathbf{Z}_i$ in (3) with $\int_0^{t_0} \boldsymbol{\beta}(u)\mathbf{Z}_i(u)dw_i(u)$ and parametrizing $\boldsymbol{\beta}(u)$ via basis function expansions, where $w_i(u)$ is a given weight function and $\mathbf{Z}_i(u)$ is the covariate level at time u. Note that only covariate information collected *up to* t_0 would be used to predict the residual life status $D_{\mathbb{L}}^{t_0+\tau} = I(T_{\mathbb{L}} \leq \tau + t_0) = I(R_{\mathbb{L}}^{t_0} \leq \tau)$. This differs from the standard Cox model with time-varying covariates, as in Kalbfleisch and Prentice (2002) (Eq. 6.14) in which covariate information collected up to time t is used to estimate the instantaneous hazard at time t. Details on building prediction tools and evaluation model consistency in this setting can be found in Jewell and Nielsen (1993).

When the underlying patient population consists of a mixture of short term survivors and long term survivors, the optimal prediction score for short term outcomes may be different from that of long term outcomes. To provide a comprehensive system for prediction, one may develop prognostic rules for short term survival based on robust procedures such as fitting a truncated Cox model as in Cai et al. (2010) or time-specific generalized linear models as in Zheng et al. (2006); Uno et al. (2007). Subsequently, the conditional prognosis rules can be developed by fitting the proposed landmark models. Such time-specific rules are likely to yield linear scores with higher accuracy compared to those obtained by fitting a global Cox model. For example, under the normal mixture configuration used in the simulation study, the ROC curves for predicting t_0 survival using the truncated Cox model vs. the global Cox model and for predicting $t_0 + \tau$ survival using the landmark model vs. the global Cox model are shown in Figure 1. The results show that both the truncated Cox model and the landmark model give better prediction rules for t_0 and $t_0 + \tau$ survival, respectively.

When $T_{\mathbb{L}} = T_{\mathbb{S}} = T$ and T follows a Cox model, our proposed procedure remains valid but is less efficient in estimating the regression coefficients when compared to those based on a global Cox model. However, the efficiency loss is minimal for the estimated accuracy measures. In practice, it is important to assess the validity of the global model and determine whether a common risk score should be used for the prediction of both short term and long

term outcomes. When the number of markers available for combination is not small relative to the number of observed events, we recommend to estimate the accuracy measures based on the cross-validation and construct confidence intervals by centering at the cross-validated point estimate but with width determined by the resampling procedure for the apparent error.

Appendix

Throughout, we assume that the joint density of $(T_{\mathbb{L}}, T_{\mathbb{S}}, \mathbf{Z})$ is twice continuously differentiable, $\mathbf{Z}$ are bounded, and $\text{pr}(X_{\mathbb{L}} > t_0 + \tau, X_{\mathbb{S}} > t_0) > 0$, C is independent of $(T_{\mathbb{L}}, T_{\mathbb{S}}, \mathbf{Z})$ with a survival function $G(\cdot)$.

Appendix A: Consistency and Large Sample Properties of $\widehat{\beta}$

To establish the convergence of $\widehat{\boldsymbol{\beta}}_{t_0}$ under possible model mis-specification, let $\widehat{S}^{(k)}(t, \boldsymbol{\beta}) = n^{-1} \sum_{i=1}^{n} \exp\{\boldsymbol{\beta}^{\mathsf{T}} \mathbf{Z}_i\} Z_i^{\otimes k} I(X_{\mathbb{L}i} \geq t, X_{\mathbb{S}i} > t_0)$, $S^{(k)}(t, \boldsymbol{\beta}) = E\{\widehat{S}^{(k)}(t, \boldsymbol{\beta})\}$ and

$$\ell_{t_0}(\boldsymbol{\beta}) = E \int_{t_0}^{\infty} \left[I(X_{\mathbb{L}i} \geq t_0) \left\{ \boldsymbol{\beta}^{\mathsf{T}} \mathbf{Z}_i - \log S^{(0)}(t, \boldsymbol{\beta}) \right\} dN_i(s) \right],$$

where and for any vector $\boldsymbol{a}$, $\boldsymbol{a}^{\otimes 0} = 1$, $\boldsymbol{a}^{\otimes 1} = \boldsymbol{a}$ and $\boldsymbol{a}^{\otimes 2} = \boldsymbol{a}\boldsymbol{a}^{\mathsf{T}}$. It follows from similar arguments as given by Hjort (1992) that $\ell_{t_0}(\boldsymbol{\beta})$ is a concave function of $\boldsymbol{\beta}$ and thus $\ell_{t_0}(\boldsymbol{\beta})$ has a unique maximizer, denoted by $\boldsymbol{\beta}_{t_0}$. We assume that $\boldsymbol{\beta}_{t_0}$ is an interior point of a compact parameter space. Without censoring, it can be shown using a penalized quasi-likelihood approximation (Breslow and Clayton, 1993) that in the neighborhood of β_{t_0}, $\ell_{t_0}(\boldsymbol{\beta})$ can be approximated by the covariance between the linear score and the survival status. In view of Theorem 2.1 of Newey et al. (1994), to show that $\widehat{\boldsymbol{\beta}}_{t_0}$ is a consistent estimator of $\boldsymbol{\beta}_{t_0}$, it suffices to show that $\widehat{\ell}_{t_0}(\boldsymbol{\beta})$ converges to $\ell_{t_0}(\boldsymbol{\beta})$ uniformly in $\boldsymbol{\beta}$. It follows from a uniform law of large numbers (ULLN) (Pollard, 1990) that $\sup_{t,\boldsymbol{\beta}} |\widehat{S}^{(k)}(t, \boldsymbol{\beta}) - S^{(k)}(t, \boldsymbol{\beta})| \to 0$, in probability. This, together with another application of a ULLN, implies that $\widehat{\ell}_{t_0}(\boldsymbol{\beta})$ converges to $\ell_{t_0}(\boldsymbol{\beta})$ uniformly in $\boldsymbol{\beta}$. Therefore, $\widehat{\beta}_{t_0} \to \boldsymbol{\beta}_{t_0}$ in probability regardless the adequacy of model (2).

To derive the limiting distribution of $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{t_0} - \boldsymbol{\beta}_{t_0})$, we take a Taylor series expansion of the score function

$$\widehat{\mathbf{U}}(\boldsymbol{\beta}) = \frac{\partial \widehat{\ell}_{t_0}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = n^{-1} \sum_{i=1}^{n} \int_{t_0}^{\infty} \left\{ \mathbf{Z}_i - \frac{\widehat{S}^{(1)}(t, \boldsymbol{\beta})}{\widehat{S}^{(0)}(t, \boldsymbol{\beta})} \right\} dN_i(t) I(X_{\mathbb{S}i} > t_0)$$

and obtain

$$0 = n^{\frac{1}{2}}\widehat{\mathbf{U}}(\boldsymbol{\beta}_{t_0}) + \widehat{\mathbb{A}}(\widehat{\boldsymbol{\beta}}^*_{t_0})n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{t_0} - \boldsymbol{\beta}_{t_0}),$$

where $\widehat{\mathbb{A}}(\boldsymbol{\beta}) = \partial\widehat{\mathbf{U}}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}^{\mathsf{T}}$ and $\widehat{\boldsymbol{\beta}}^*_{t_0}$ satisfies $\|\widehat{\boldsymbol{\beta}}^*_{t_0} - \boldsymbol{\beta}_{t_0}\| \leq \|\widehat{\boldsymbol{\beta}}_{t_0} - \boldsymbol{\beta}_{t_0}\|$. First, it follows from the uniform convergence of $\widehat{S}^{(k)}(t,\boldsymbol{\beta})$ and a ULLN that $\widehat{\mathbb{A}}(\boldsymbol{\beta}) \to \mathbb{A}(\boldsymbol{\beta})$ in probability uniformly in $\boldsymbol{\beta}$, where $\mathbb{A}(\boldsymbol{\beta}) = \partial^2\ell_{t_0}(\boldsymbol{\beta})/\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^{\mathsf{T}}$. This, together with the consistency of $\widehat{\boldsymbol{\beta}}_{t_0}$ implies that $\widehat{\mathbb{A}}(\widehat{\boldsymbol{\beta}}^*_{t_0}) \to \mathbb{A}(\boldsymbol{\beta}_{t_0})$ in probability.

We next derive an asymptotic expansion for $n^{\frac{1}{2}}\widehat{\mathbf{U}}(\boldsymbol{\beta}_{t_0})$. To this end, we let $\widehat{\xi}(t) = n^{-1}\sum_{i=1}^n N_i(t)I(X_{\mathbb{S}i} > t_0)$ and $\xi(t) = E\{\widehat{\xi}(t)\}$ and write

$$n^{\frac{1}{2}}\widehat{\mathbf{U}}(\boldsymbol{\beta}) \;=\; n^{-\frac{1}{2}}\sum_{X_{\mathbb{S}i}>t_0}\left[\int_{t_0}^{\infty}\left\{\mathbf{Z}_i - \frac{S^{(1)}(t,\boldsymbol{\beta})}{S^{(0)}(t,\boldsymbol{\beta})}\right\}dN_i(t)\right] + n^{\frac{1}{2}}\int_{t_0}^{\infty}\left\{\frac{S^{(1)}(t,\boldsymbol{\beta})}{S^{(0)}(t,\boldsymbol{\beta})} - \frac{\hat{S}^{(1)}(t,\boldsymbol{\beta})}{\hat{S}^{(0)}(t,\boldsymbol{\beta})}\right\}d\widehat{\xi}(t).$$

By a Functional Central Limit Theorem (FCLT) (Pollard, 1990) and the uniform consistency of $\widehat{S}^{(k)}(t,\boldsymbol{\beta})$,

$$n^{\frac{1}{2}}\left\{\frac{S^{(1)}(t,\boldsymbol{\beta})}{S^{(0)}(t,\boldsymbol{\beta})} - \frac{\hat{S}^{(1)}(t,\boldsymbol{\beta})}{\hat{S}^{(0)}(t,\boldsymbol{\beta})}\right\} \approx -n^{-\frac{1}{2}}\sum_{j=1}^{n}\frac{I(X_{\mathbb{L}j} \geq t, X_{\mathbb{S}j} > t_0)\exp\{\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_j\}}{S^{(0)}(t,\boldsymbol{\beta})}\left\{\mathbf{Z}_j - \frac{S^{(1)}(t,\boldsymbol{\beta})}{S^{(0)}(t,\boldsymbol{\beta})}\right\} \tag{4}$$

Moreover, by a ULLN, $\sup_t|\widehat{\xi}(t) - \xi(t)| \to 0$ in probability. This, together with (4), a strong representation theorem (Pollard (1990)) and Lemma A.3 of Bilias et al. (1997), implies that

$$n^{\frac{1}{2}}\int_{t_0}^{\infty}\left\{\frac{S^{(1)}(t,\boldsymbol{\beta})}{S^{(0)}(t,\boldsymbol{\beta})} - \frac{\hat{S}^{(1)}(t,\boldsymbol{\beta})}{\hat{S}^{(0)}(t,\boldsymbol{\beta})}\right\}d\widehat{\xi}(t) = n^{\frac{1}{2}}\int_{t_0}^{\infty}\left\{\frac{S^{(1)}(t,\boldsymbol{\beta})}{S^{(0)}(t,\boldsymbol{\beta})} - \frac{\hat{S}^{(1)}(t,\boldsymbol{\beta})}{\hat{S}^{(0)}(t,\boldsymbol{\beta})}\right\}d\xi(t) + o_p(1).$$

This, together with the convergence of $\widehat{\mathbb{A}}$, implies that

$$n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{t_0} - \boldsymbol{\beta}_{t_0}) \approx n^{-\frac{1}{2}}\sum_{i=1}^{n}\mathbf{U}_i(\boldsymbol{\beta}_{t_0}). \tag{5}$$

where

$$\mathbf{U}_i(\boldsymbol{\beta}) = -\mathbb{A}(\boldsymbol{\beta}_{t_0})^{-1}\int_{t_0}^{\infty}I(X_{\mathbb{S}i} > t_0)\left\{\mathbf{Z}_i - \frac{S^{(1)}(t,\boldsymbol{\beta})}{S^{(0)}(t,\boldsymbol{\beta})}\right\}\left\{dN_i(t) - \frac{I(X_{\mathbb{L}i} \geq t)\exp\{\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_i\}}{S^{(0)}(t,\boldsymbol{\beta})}d\xi(t)\right\}.$$

By a central limit theorem, $n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}}_{t_0} - \boldsymbol{\beta}_{t_0})$ converges in distribution to a multivariate normal with mean 0 and covariance matrix $E\{\mathbf{U}_i(\boldsymbol{\beta}_{t_0})\mathbf{U}_i(\boldsymbol{\beta}_{t_0})^{\mathsf{T}}\}$.

Appendix B: Consistency and Large Sample Properties of accuracy measures

In this section, we derive large sample properties for the estimated accuracy measures. We only provide details for the estimated sensitivity function, but note that the same

arguments can be used for the estimated specificity, PPV and NPV functions. To this end, we let $H(\boldsymbol{\beta}, c) = \mathrm{pr}_{\Omega_{t_0}}(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z} > c \mid D_{\mathbb{L}}^t = 1)$, $W_i = I(X_{\mathbb{L}i} > t_0 + \tau)/G(t_0 + \tau) + I(X_{\mathbb{L}i} \leq t_0 + \tau)1/G(X_{\mathbb{L}i})$,

$$\widehat{H}(\boldsymbol{\beta}, c) = \frac{\sum_{i=1}^n \widehat{W}_i I(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_i \geq c) I(X_{\mathbb{L}i} \leq t_0 + \tau, X_{\mathbb{S}i} > t_0)}{\sum_{i=1}^n \widehat{W}_i I(X_{\mathbb{L}i} \leq t_0 + \tau, X_{\mathbb{S}i} > t_0)}$$

$$\text{and} \qquad \widetilde{H}(\boldsymbol{\beta}, c) = \frac{\sum_{i=1}^n W_i I(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_i \geq c) I(X_{\mathbb{L}i} \leq t_0 + \tau, X_{\mathbb{S}i} > t_0)}{\sum_{i=1}^n W_i I(X_{\mathbb{L}i} \leq t_0 + \tau, X_{\mathbb{S}i} > t_0)}$$

To establish the uniform consistency of the estimated sensitivity function, i.e. $\sup_{c\in[c_l,c_r]} |\widehat{H}(\widehat{\boldsymbol{\beta}}_{t_0}, c) - H(\boldsymbol{\beta}_{t_0}, c)| \to 0$ in probability, we first show that $\sup_{\boldsymbol{\beta},c} |\widehat{H}(\boldsymbol{\beta}, c) - \widetilde{H}(\boldsymbol{\beta}, c)| \to 0$ in probability, where $1 > H(\boldsymbol{\beta}_{t_0}, c_l) > H(\boldsymbol{\beta}_{t_0}, c_r) > 0$. Here and in the sequel, the sup or inf is taken over $\Omega_{\boldsymbol{\beta}_{t_0}}^{(n)} = \{\boldsymbol{\beta} : \boldsymbol{\beta} = \boldsymbol{\beta}_{t_0} + O_p(n^{-\frac{1}{2}})\}$ for $\boldsymbol{\beta}$ and over $[c_l, c_r]$ for c. It is straightforward to see that

$$|\widehat{H}(\boldsymbol{\beta}, c) - \widetilde{H}(\boldsymbol{\beta}, c)| \leq \frac{|\widehat{\mathcal{N}}(\boldsymbol{\beta}, c) - \widetilde{\mathcal{N}}(\boldsymbol{\beta}, c)|}{\widehat{\mathcal{D}}(\boldsymbol{\beta}, c)} + \frac{|\widehat{\mathcal{D}}(\boldsymbol{\beta}, c) - \widetilde{\mathcal{D}}(\boldsymbol{\beta}, c)|}{G(t_0 + \tau)\widehat{\mathcal{D}}(\boldsymbol{\beta}, c)\widetilde{\mathcal{D}}(\boldsymbol{\beta}, c)}$$

where

$$\widetilde{\mathcal{N}}(\boldsymbol{\beta}, c) = n^{-1}\sum_{i=1}^n W_i I(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_i \geq c) I(X_{\mathbb{L}i} \leq t_0 + \tau, X_{\mathbb{S}i} > t_0), \quad \widetilde{\mathcal{D}}(\boldsymbol{\beta}, c) = n^{-1}\sum_{i=1}^n W_i I(X_{\mathbb{L}i} \leq t_0 + \tau, X_{\mathbb{S}i} > t_0),$$

$$\widehat{\mathcal{N}}(\boldsymbol{\beta}, c) = n^{-1}\sum_{i=1}^n \widehat{W}_i I(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_i \geq c) I(X_{\mathbb{L}i} \leq t_0 + \tau, X_{\mathbb{S}i} > t_0), \quad \widehat{\mathcal{D}}(\boldsymbol{\beta}, c) = n^{-1}\sum_{i=1}^n \widehat{W}_i I(X_{\mathbb{L}i} \leq t_0 + \tau, X_{\mathbb{S}i} > t_0)$$

Furthermore,

$$\widehat{\mathcal{N}}(\boldsymbol{\beta}, c) - \widetilde{\mathcal{N}}(\boldsymbol{\beta}, c) = \int \left[\frac{1}{\widehat{G}(s)} - \frac{1}{G(s)}\right] d\left\{n^{-1}\sum_{i=1}^n I(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_i \geq c, t_0 < X_{\mathbb{L}i} \leq s, X_{\mathbb{S}i} > t_0, \delta_{\mathbb{L}i} = 1)\right\}$$

It follows from the uniform consistency of $\widehat{G}(\cdot)$ (Fleming and Harrington, 1991), a uniform law of large numbers, along with Lemma A.3 of Bilias et al. (1997) that $\sup_{\boldsymbol{\beta},c} |\widehat{\mathcal{N}}(\boldsymbol{\beta}, c) - \widetilde{\mathcal{N}}(\boldsymbol{\beta}, c)| \to 0$ in probability. Similarly, $\sup_{\boldsymbol{\beta},c} |\widehat{\mathcal{D}}(\boldsymbol{\beta}, c) - \widetilde{\mathcal{D}}(\boldsymbol{\beta}, c)| \to 0$ in probability. This, together with $\sup_{\boldsymbol{\beta},c} |\widetilde{\mathcal{D}}(\boldsymbol{\beta}, c) - E\{\widetilde{\mathcal{D}}(\boldsymbol{\beta}, c)\}| \to 0$ in probability by a ULLN and and $\inf_{\boldsymbol{\beta},c} E\{\widetilde{\mathcal{D}}(\boldsymbol{\beta}, c)\} > 0$, implies the uniform in probability convergence of $\widehat{H}(\boldsymbol{\beta}, c) - \widetilde{H}(\boldsymbol{\beta}, c) \to 0$. On the other hand, by a ULLN, $\sup_{\boldsymbol{\beta},c} |\widetilde{H}(\boldsymbol{\beta}, c) - H(\boldsymbol{\beta}, c)| \to 0$, in probability. This, together with the consistency of $\widehat{\boldsymbol{\beta}}_{t_0}$, implies the uniform consistency of $\widehat{\mathrm{Sens}}_{t_0,\tau}(c) = \widehat{H}(\widehat{\boldsymbol{\beta}}_{t_0}, c)$ for $\mathrm{Sens}_{t_0,\tau}(c) = H(\boldsymbol{\beta}_{t_0}, c)$.

We now approximate the distribution of $\widehat{\mathcal{W}}(c) = n^{\frac{1}{2}}\{\widehat{H}(\widehat{\boldsymbol{\beta}}_{t_0}, c) - H(\boldsymbol{\beta}_{t_0}, c)\} = n^{\frac{1}{2}}\{q_0\widehat{q}_1(\widehat{\boldsymbol{\beta}}_{t_0}, \widehat{G}) - q_1(\boldsymbol{\beta}_{t_0})\widehat{q}_0(\widehat{G})\}/\{q_0\widehat{q}_0(\widehat{G})\}$, where $q_0 = \mathrm{pr}(X_{\mathbb{L}i} \leq t_0 + \tau, X_{\mathbb{S}i} > t_0)$, $\widehat{q}_1(\boldsymbol{\beta}, G) = n^{-1}\sum_{i=1}^n \delta_{\mathbb{L}i} I(\boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_i \geq c, X_{\mathbb{L}i} \leq t_0 + \tau, X_{\mathbb{S}i} > t_0)/G(X_{\mathbb{L}i})$, $q_1(\boldsymbol{\beta}) = \mathrm{pr}(X_{\mathbb{L}i} \leq t_0 + \tau, X_{\mathbb{S}i} > t_0, \boldsymbol{\beta}^{\mathsf{T}}\mathbf{Z}_i \geq c)$ and

$\widehat{q}_0(G) = n^{-1}\sum_{i=1}^n \delta_{\mathbb{L}i} I(X_{\mathbb{L}i} \leq t_0+\tau, X_{\mathbb{S}i} > t_0)/G(X_{\mathbb{L}i})$. We begin by examining the numerator and write $n^{\frac{1}{2}}\{q_0\widehat{q}_1(\widehat{\boldsymbol{\beta}}_{t_0}, \widehat{G}) - q_1(\boldsymbol{\beta}_{t_0})\widehat{q}_0(\widehat{G})\} = (B_1)+(B_2)+(B_3)+(B_4)$, where

$$(B_1) = n^{\frac{1}{2}}\left[\widehat{q}_1(\boldsymbol{\beta}_{t_0}, G)q_0 - \widehat{q}_0(G)q_1(\boldsymbol{\beta}_{t_0})\right], \quad (B_2) = n^{\frac{1}{2}}\left[\widehat{q}_1(\widehat{\boldsymbol{\beta}}_{t_0}, \widehat{G}) - \widehat{q}_1(\boldsymbol{\beta}_{t_0}, \widehat{G})\right] q_0,$$
$$(B_3) = -n^{\frac{1}{2}}\left[\widehat{q}_0(\widehat{G}) - \widehat{q}_0(G)\right] q_1(\boldsymbol{\beta}_{t_0}), \qquad (B_4) = n^{\frac{1}{2}}\left[\widehat{q}_1(\boldsymbol{\beta}_{t_0}, \widehat{G}) - \widehat{q}_1(\boldsymbol{\beta}_{t_0}, G)\right] q_0.$$

It is straightforward to show that $(B_1) = n^{-\frac{1}{2}}\sum_{i=1}^n B_{1i}(c)$, where $B_{1i}(c) = \delta_{\mathbb{L}i} I(X_{\mathbb{L}i} \leq t_0+\tau, X_{\mathbb{S}i} > t_0)\{I(\boldsymbol{\beta}_{t_0}^{\intercal}\mathbf{Z}_i \geq c)q_0 - q_1(\boldsymbol{\beta}_{t_0})\}/G(X_{\mathbb{L}i})$. For (B_2), we write $\widehat{q}_1(\boldsymbol{\beta}, \widehat{G}) = \int_{t_0}^{t_0+\tau} \widehat{\eta}_1(dt, c, \boldsymbol{\beta})/\widehat{G}(t)$, where $\widehat{\eta}_1(t, c, \boldsymbol{\beta}) = n^{-1}\sum_{i=1}^n I(X_{\mathbb{L}i} \leq t, X_{\mathbb{S}i} > t_0) I(\boldsymbol{\beta}_{t_0}^{\intercal}\mathbf{Z}_i \geq c)\delta_{\mathbb{L}i}$. It follows from a FCLT that $n^{\frac{1}{2}}(\widehat{\eta}_1(t, c, \boldsymbol{\beta}) - \eta_1(t, c, \boldsymbol{\beta}))$ converges weakly to a zero-mean Gaussian Process in $(t, c, \boldsymbol{\beta})$ and thus is equicontinuous, where $\eta_1(t, c, \boldsymbol{\beta}) = E\{\widehat{\eta}_1(t, c, \boldsymbol{\beta})\}$. This, together with the uniform consistency of $\widehat{G}$, Lemma A.3 of Bilias et al. (1997), and a Taylor series expansion, implies that $(B_2) \approx q_0 n^{\frac{1}{2}}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{t_0})^{\intercal}\int_{t_0}^{t_0+\tau} G(t)^{-1}\dot{\boldsymbol{\eta}}_1(dt, c, \boldsymbol{\beta}_{t_0})$, where $\dot{\boldsymbol{\eta}}_1(t, c, \boldsymbol{\beta}) = \partial\eta_1(t, c, \boldsymbol{\beta})/\partial\boldsymbol{\beta}$. This, together with (5), implies that $(B_2) \approx n^{-\frac{1}{2}}\sum_{i=1}^n B_{2i}$, where $B_{2i}(c) = q_0\int_{t_0}^{t_0+\tau} G(t)^{-1}\dot{\boldsymbol{\eta}}_1(dt, c, \boldsymbol{\beta}_{t_0})^{\intercal}\mathbf{U}_i(\boldsymbol{\beta}_{t_0})$.

To account for the variability due to $\widehat{G}$ in (B_3), we first note that $n^{\frac{1}{2}}\{G(t)/\widehat{G}(t) - 1\} \approx n^{-\frac{1}{2}}\sum_{i=1}^n U_{Gi}(t)$ (Fleming and Harrington, 1991), where $U_{Gi}(t) = \int_0^t dM_{Ci}(s)/\pi_{\mathbb{S}}(s)$, $\pi_{\mathbb{S}}(s) = \text{pr}(X_{\mathbb{L}i} > s)$, $M_{Ci}(s) = I(X_{\mathbb{L}i} \leq s, \delta_{\mathbb{L}i} = 0) + \int_0^t I(X_{\mathbb{L}j} > s) d\log\{G(s)\}$. This, together with a Strong Representation Theorem (Pollard, 1990) and Lemma A.3 of Bilias et al. (1997), implies that $(B_3) \approx q_1(\boldsymbol{\beta}_{t_0})\int_{t_0}^{t_0+\tau} n^{\frac{1}{2}}\{G(t)/\widehat{G}(t) - 1\}\text{pr}(T_{\mathbb{L}} \leq dt, X_{\mathbb{S}} > t_0) \approx n^{-\frac{1}{2}}\sum_{i=1}^n B_{3i}$, where $B_{3i} = q_1(\boldsymbol{\beta}_{t_0})\int_{t_0}^{t_0+\tau} U_{Gi}(t)\text{pr}(T_{\mathbb{L}} \leq dt, X_{\mathbb{S}} > t_0)$. Similarly, we have $(B_4) = n^{-\frac{1}{2}}n^{-1}\sum_{i=1}^n B_{4i}(c)$, where $B_{4i}(c) = q_0\int_{t_0}^{t_0+\tau} U_{Gi}(t)\text{pr}(T_{\mathbb{L}} \leq dt, X_{\mathbb{S}} > t_0, \boldsymbol{\beta}_{t_0}^{\intercal}\mathbf{Z} > c)$.

Combining the above expansions for $(B_1), (B_2), (B_3)$ and (B_4), we have

$$\widehat{\mathcal{W}}(c) \approx \{q_0\widehat{q}_0(\widehat{G})\} - 1 n^{-\frac{1}{2}}\sum_{i=1}^n \{B_{1i}(c) + B_{2i}(c) + B_{3i} + B_{4i}(c)\}.$$

On the other hand, it follows from the uniform consistency of $\widehat{G}$, a ULLN Lemma A.3 of Bilias et al. (1997), that $\widehat{q}_0(\widehat{G}) \to q_0$. Therefore, $\widehat{\mathcal{W}}(c) \approx n^{-\frac{1}{2}}\sum_{i=1}^n U_{\text{Sens}_i}(c)$, where $U_{\text{Sens}_i}(c) = q_0^{-2}\{B_{1i}(c) + B_{2i}(c) + B_{3i} + B_{4i}(c)\}$. It then follows from a FCLT that $\widehat{\mathcal{W}}(c)$ converges weakly to a mean-zero Gaussian process with covariance function $\text{cov}\{U_{\text{Sens}_i}(c), U_{\text{Sens}_i}(c')\}$.

Table 1. *Regression parameter and accuracy measure estimates under the log-normal mixture model for T and exponential model for C using the proposed landmark method (Land) vs. the global Cox method (Cox), with corresponding empirical standard errors (ESE), average of the standard error estimates from the perturbation-resampling method (ASE), and empirical coverage levels (Coverage)*

(a) Regression Coefficients

n=200										
	Truth		Average		ESE		ASE		Coverage	
	Land	Cox	Land	Cox	Land	Cox	Land	Cox	Land	Cox
β_1	0.412	-0.184	0.433	-0.173	0.352	0.265	0.331	0.250	0.928	0.944
β_2	0.047	-0.408	0.060	-0.434	0.143	0.120	0.132	0.111	0.936	0.939
β_3	-0.381	-0.444	-0.369	-0.448	0.156	0.118	0.142	0.113	0.934	0.944
β_4	-0.733	-0.676	-0.729	-0.669	0.196	0.132	0.178	0.136	0.927	0.957
β_5	-0.382	-0.381	-0.375	-0.365	0.413	0.316	0.411	0.324	0.951	0.961
n=500										
β_1	0.412	-0.184	0.405	-0.176	0.205	0.157	0.202	0.155	0.950	0.936
β_2	0.047	-0.408	0.052	-0.425	0.084	0.070	0.081	0.069	0.943	0.941
β_3	-0.381	-0.444	-0.379	-0.449	0.090	0.072	0.087	0.069	0.941	0.951
β_4	-0.733	-0.676	-0.732	-0.659	0.113	0.086	0.109	0.083	0.928	0.946
β_5	-0.382	-0.381	-0.392	-0.390	0.256	0.202	0.248	0.199	0.943	0.945
n=1000										
β_1	0.412	-0.184	0.413	-0.174	0.140	0.108	0.141	0.108	0.943	0.949
β_2	0.047	-0.408	0.051	-0.414	0.058	0.048	0.057	0.048	0.937	0.944
β_3	-0.381	-0.444	-0.380	-0.445	0.062	0.048	0.061	0.049	0.940	0.958
β_4	-0.733	-0.676	-0.728	-0.675	0.079	0.059	0.076	0.058	0.946	0.946
β_5	-0.382	-0.381	-0.391	-0.381	0.181	0.141	0.174	0.139	0.939	0.943

(b) Accuracy Measures

n=200										
	Truth		Average		ESE		ASE		Coverage	
	Land	Cox	Land	Cox	Land	Cox	Land	Cox	Land	Cox
AUC	0.741	0.668	0.756	0.675	0.051	0.067	0.048	0.064	0.891	0.920
Spec	0.319	0.207	0.360	0.243	0.119	0.106	0.129	0.117	0.952	0.960
NPV	0.771	0.686	0.790	0.707	0.074	0.114	0.078	0.120	0.885	0.892
PPV	0.557	0.519	0.577	0.535	0.067	0.059	0.072	0.065	0.952	0.962
n=500										
AUC	0.741	0.668	0.746	0.671	0.031	0.041	0.031	0.041	0.937	0.946
Spec	0.319	0.207	0.336	0.223	0.074	0.065	0.081	0.071	0.954	0.960
NPV	0.771	0.686	0.779	0.696	0.046	0.071	0.050	0.075	0.931	0.931
PPV	0.557	0.519	0.565	0.526	0.042	0.038	0.044	0.040	0.958	0.960
n=1000										
AUC	0.741	0.668	0.744	0.671	0.023	0.030	0.022	0.029	0.942	0.937
Spec	0.319	0.207	0.329	0.216	0.055	0.046	0.057	0.049	0.958	0.958
NPV	0.771	0.686	0.775	0.691	0.034	0.051	0.035	0.052	0.922	0.923
PPV	0.557	0.519	0.561	0.522	0.030	0.026	0.031	0.027	0.957	0.963

Table 2. *Regression parameter and accuracy measure estimates under the log-normal mixture model for T and exponential mixture model for C using the proposed landmark method (Land) vs. the global Cox method (Cox), with corresponding empirical standard errors (ESE), average of the standard error estimates from the perturbation-resampling method (ASE), and empirical coverage levels (Coverage)*

(a) Regression Coefficients

n=200										
	Truth		Average		ESE		ASE		Coverage	
	Land	Cox	Land	Cox	Land	Cox	Land	Cox	Land	Cox
β_1	0.320	-0.129	0.344	-0.104	0.285	0.229	0.274	0.221	0.943	0.939
β_2	0.041	-0.298	0.039	-0.346	0.123	0.102	0.111	0.097	0.921	0.935
β_3	-0.392	-0.444	-0.392	-0.448	0.127	0.103	0.120	0.099	0.933	0.943
β_4	-0.764	-0.735	-0.770	-0.734	0.168	0.126	0.151	0.122	0.922	0.950
β_5	-0.397	-0.396	-0.366	-0.361	0.371	0.293	0.345	0.285	0.933	0.947
n=500										
β_1	0.320	-0.129	0.328	-0.121	0.177	0.138	0.169	0.137	0.936	0.946
β_2	0.041	-0.298	0.046	-0.313	0.077	0.060	0.069	0.060	0.916	0.942
β_3	-0.392	-0.444	-0.387	-0.445	0.078	0.064	0.074	0.061	0.931	0.938
β_4	-0.764	-0.735	-0.770	-0.734	0.098	0.079	0.093	0.075	0.937	0.926
β_5	-0.397	-0.396	-0.384	-0.388	0.219	0.180	0.211	0.175	0.942	0.938
n=1000										
β_1	0.320	-0.129	0.320	-0.125	0.118	0.096	0.119	0.096	0.959	0.948
β_2	0.041	-0.298	0.044	-0.307	0.052	0.043	0.049	0.042	0.937	0.946
β_3	-0.392	-0.444	-0.391	-0.444	0.053	0.043	0.052	0.043	0.944	0.942
β_4	-0.764	-0.735	-0.764	-0.733	0.070	0.054	0.066	0.053	0.926	0.942
β_5	-0.397	-0.396	-0.398	-0.394	0.149	0.124	0.149	0.123	0.952	0.942

(b) Accuracy Measures

n=200										
	Truth		Average		ESE		ASE		Coverage	
	Land	Cox	Land	Cox	Land	Cox	Land	Cox	Land	Cox
AUC	0.739	0.691	0.749	0.694	0.046	0.055	0.045	0.056	0.923	0.948
Spec	0.311	0.223	0.342	0.254	0.110	0.098	0.119	0.112	0.952	0.973
NPV	0.766	0.701	0.780	0.719	0.072	0.097	0.077	0.108	0.918	0.916
PPV	0.553	0.523	0.571	0.539	0.062	0.055	0.066	0.061	0.949	0.958
n=500										
AUC	0.739	0.691	0.743	0.692	0.030	0.036	0.029	0.036	0.932	0.939
Spec	0.311	0.223	0.322	0.232	0.070	0.062	0.075	0.069	0.959	0.962
NPV	0.766	0.701	0.773	0.707	0.045	0.063	0.049	0.069	0.933	0.939
PPV	0.553	0.523	0.558	0.526	0.040	0.036	0.041	0.038	0.948	0.959
n=1000										
AUC	0.739	0.691	0.742	0.692	0.021	0.025	0.021	0.025	0.944	0.949
Spec	0.311	0.223	0.317	0.231	0.052	0.045	0.053	0.048	0.945	0.955
NPV	0.766	0.701	0.769	0.706	0.033	0.045	0.034	0.048	0.927	0.938
PPV	0.553	0.523	0.557	0.527	0.029	0.026	0.029	0.026	0.947	0.948

Table 3. *Regression parameter and accuracy measure estimates under the Cox model for T using the proposed landmark method (Land) vs. the global Cox method (Cox), with corresponding empirical standard errors (ESE), average of the standard error estimates from the perturbation-resampling method (ASE), and empirical coverage levels (Coverage)*

n=200										
	Truth		Average		ESE		ASE		Coverage	
	Land	Cox	Land	Cox	Land	Cox	Land	Cox	Land	Cox
β_1	-0.603	-0.603	-0.605	-0.605	0.296	0.271	0.283	0.257	0.932	0.931
β_2	-0.301	-0.302	-0.301	-0.302	0.137	0.121	0.132	0.115	0.940	0.937
β_3	-0.301	-0.302	-0.298	-0.298	0.135	0.115	0.131	0.114	0.940	0.946
β_4	-0.603	-0.603	-0.603	-0.604	0.181	0.157	0.172	0.147	0.935	0.927
β_5	-0.302	-0.302	-0.301	-0.299	0.379	0.330	0.361	0.319	0.935	0.928
n=500										
β_1	-0.603	-0.603	-0.604	-0.604	0.182	0.166	0.172	0.158	0.931	0.944
β_2	-0.301	-0.302	-0.302	-0.303	0.083	0.075	0.080	0.071	0.937	0.930
β_3	-0.301	-0.302	-0.301	-0.301	0.084	0.074	0.080	0.070	0.939	0.931
β_4	-0.603	-0.603	-0.604	-0.603	0.113	0.097	0.105	0.090	0.925	0.921
β_5	-0.302	-0.302	-0.298	-0.299	0.231	0.202	0.221	0.197	0.936	0.947
n=1000										
β_1	-0.603	-0.603	-0.605	-0.604	0.121	0.113	0.121	0.111	0.946	0.941
β_2	-0.301	-0.302	-0.302	-0.302	0.057	0.051	0.056	0.050	0.948	0.942
β_3	-0.301	-0.302	-0.300	-0.300	0.058	0.051	0.056	0.050	0.947	0.950
β_4	-0.603	-0.603	-0.602	-0.602	0.077	0.066	0.074	0.063	0.936	0.932
β_5	-0.302	-0.302	-0.301	-0.301	0.159	0.142	0.156	0.139	0.937	0.941

(b) Accuracy Measures

n=200										
	Truth		Average		ESE		ASE		Coverage	
	Land	Cox	Land	Cox	Land	Cox	Land	Cox	Land	Cox
AUC	0.865	0.865	0.870	0.869	0.030	0.030	0.029	0.030	0.910	0.915
Spec	0.588	0.588	0.601	0.599	0.102	0.102	0.112	0.112	0.931	0.935
NPV	0.907	0.907	0.914	0.914	0.021	0.021	0.024	0.025	0.921	0.927
PPV	0.568	0.568	0.584	0.582	0.074	0.074	0.080	0.080	0.951	0.949
n=500										
AUC	0.865	0.865	0.867	0.866	0.019	0.019	0.019	0.019	0.946	0.947
Spec	0.588	0.588	0.591	0.591	0.066	0.067	0.073	0.073	0.956	0.960
NPV	0.907	0.907	0.910	0.910	0.013	0.013	0.015	0.015	0.929	0.940
PPV	0.568	0.568	0.573	0.573	0.047	0.047	0.051	0.051	0.959	0.961
n=1000										
AUC	0.865	0.865	0.866	0.866	0.013	0.013	0.014	0.014	0.959	0.960
Spec	0.588	0.588	0.588	0.587	0.046	0.046	0.051	0.051	0.953	0.955
NPV	0.907	0.907	0.908	0.908	0.010	0.009	0.010	0.010	0.947	0.950
PPV	0.568	0.568	0.569	0.569	0.033	0.033	0.036	0.036	0.965	0.956

Table 4. *Estimates of the regression coefficient along with the accuracy of the resulting prognosis rules based on truncated Cox model for predicting $D_{\mathbb{S}}^{t_0}$ (Trunc), the global Cox model (Cox), as well as the landmark procedure (Land) for the prediction of $D_{\mathbb{L}}^{t_0+\tau} \mid T_{\mathbb{S}} > t_0$, where both $T_{\mathbb{S}}$ and $T_{\mathbb{L}}$ represent metastasis free survival. Shown also are the standard error (SE) estimates based on the proposed resampling methods. The regression coefficients are normalized such that $\|\vec{\beta}\| = 1$.*

(a) Regression Coefficients

	with gene score				without gene score			
	Estimate		SE		Estimate		SE	
$D_{\mathbb{S}}^{t_0}$	Trunc	Cox	Trunc	Cox	Trunc	Cox	Trunc	Cox
genescore	0.495	0.928	0.322	0.200				
age	-0.061	-0.093	0.028	0.020	-0.050	-0.099	0.027	0.020
grade	0.860	0.324	0.343	0.175	0.944	0.827	0.303	0.151
size	0.030	0.038	0.017	0.013	0.026	0.041	0.016	0.012
ERstatus	-0.009	0.153	0.437	0.311	-0.283	-0.520	0.363	0.272
posLN	0.102	-0.020	0.334	0.228	0.160	0.188	0.328	0.224
$D_{\mathbb{L}}^{t_0+\tau}$	Land	Cox	Land	Cox	Land	Cox	Land	Cox
genescore	0.897	0.928	0.234	0.200				
age	-0.072	-0.093	0.026	0.020	-0.225	-0.099	0.026	0.020
grade	-0.200	0.324	0.215	0.175	0.599	0.827	0.182	0.151
size	0.022	0.038	0.018	0.013	0.067	0.041	0.017	0.012
ERstatus	0.357	0.153	0.431	0.311	-0.746	-0.520	0.383	0.272
posLN	-0.147	-0.020	0.290	0.228	0.171	0.188	0.286	0.224

(b) Estimates of accuracy measures along with their standard errors (SE).

	with gene score						without gene score					
	Estimate				SE		Estimate				SE	
$D_{\mathbb{S}}^{t_0}$	Trunc		Cox		Trunc	Cox	Trunc		Cox		Trunc	Cox
	AP	CV	AP	CV			AP	CV	AP	CV		
AUC	0.770	0.733	0.754	0.728	0.035	0.057	0.761	0.730	0.755	0.730	0.039	0.057
Spec	0.451	0.396	0.428	0.375	0.109	0.112	0.456	0.400	0.433	0.390	0.156	0.101
NPV	0.961	0.955	0.959	0.956	0.011	0.020	0.961	0.945	0.959	0.948	0.032	0.037
PPV	0.243	0.239	0.235	0.234	0.049	0.051	0.244	0.242	0.236	0.238	0.058	0.039
$D_{\mathbb{L}}^{t_0+\tau}$	Land		Cox		Land	Cox	Land		Cox		Land	Cox
	AP	CV	AP	CV			AP	CV	AP	CV		
AUC	0.716	0.646	0.655	0.635	0.052	0.057	0.624	0.536	0.581	0.560	0.061	0.057
Spec	0.455	0.363	0.337	0.353	0.112	0.112	0.248	0.230	0.238	0.233	0.104	0.101
NPV	0.972	0.974	0.953	0.984	0.015	0.020	0.934	0.956	0.925	0.954	0.038	0.037
PPV	0.257	0.231	0.218	0.231	0.058	0.051	0.197	0.198	0.193	0.199	0.043	0.039

(c) Predicted proability of survival (standard errors) for the short term outcome and conditional survival for the long term outcome given the corresponding prognoses. .

	probability of $T_{\mathbb{S}} > t_0$	probability of $R_{\mathbb{L}}^{t_0} > \tau \mid T_{\mathbb{S}} > t_0$	
$\mathbb{S}_{t_0}^{+}$	0.757 (0.049)	$\mathbb{L}_{t_0}^{\tau+}$	0.736 (0.058)
		$\mathbb{L}_{t_0}^{\tau-}$	0.958 (0.079)
$\mathbb{S}_{t_0}^{-}$	0.961 (0.011)	$\mathbb{L}_{t_0}^{\tau+}$	0.837 (0.145)
		$\mathbb{L}_{t_0}^{\tau-}$	0.977 (0.020)

Table 5. *Estimated accuracy of the prognosis rules based on the global Cox model (Cox) and the landmark procedure (Land) in predicting* $D_{\mathbb{L}}^{t_0+\tau} \mid T_{\mathbb{S}} > t_0$, *where both* $T_{\mathbb{S}}$ *represents metastasis free survival and* $T_{\mathbb{L}}$ *represents overall survival. Shown also are the standard error (SE) estimates based on the proposed resampling methods.*

	with gene score						without gene score					
	Estimate				SE		Estimate				SE	
	Land		Cox		Land	Cox	Land		Cox		Land	Cox
	AP	CV	AP	CV			AP	CV	AP	CV		
AUC	0.750	0.678	0.723	0.687	0.048	0.056	0.654	0.587	0.653	0.637	0.056	0.053
Spec	0.550	0.432	0.495	0.475	0.122	0.091	0.459	0.378	0.450	0.454	0.106	0.072
NPV	0.984	1.000	0.982	1.000	0.009	0.008	0.983	0.996	0.982	0.999	0.011	0.010
PPV	0.190	0.179	0.173	0.187	0.059	0.051	0.164	0.159	0.162	0.176	0.042	0.040

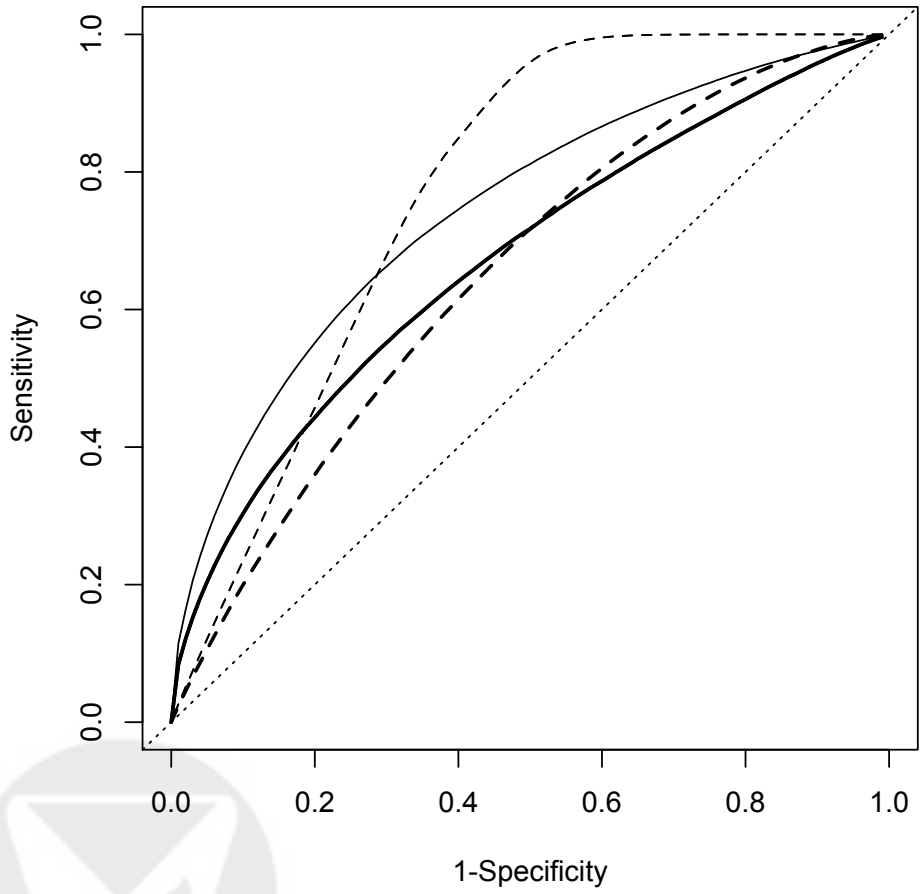

Fig. 1. *ROC curves for the prediction of (i)* $T_{\text{L}} < \tau + t_0$ *among* Ω_{t_0} *based on the landmark (solid thin line) and global Cox (solid thick line) procedures and for the prediction of (ii)* $T_{\text{S}} < \tau$ *based on the truncated (dashed thin line) and global Cox (dashed thick line) procedures.*

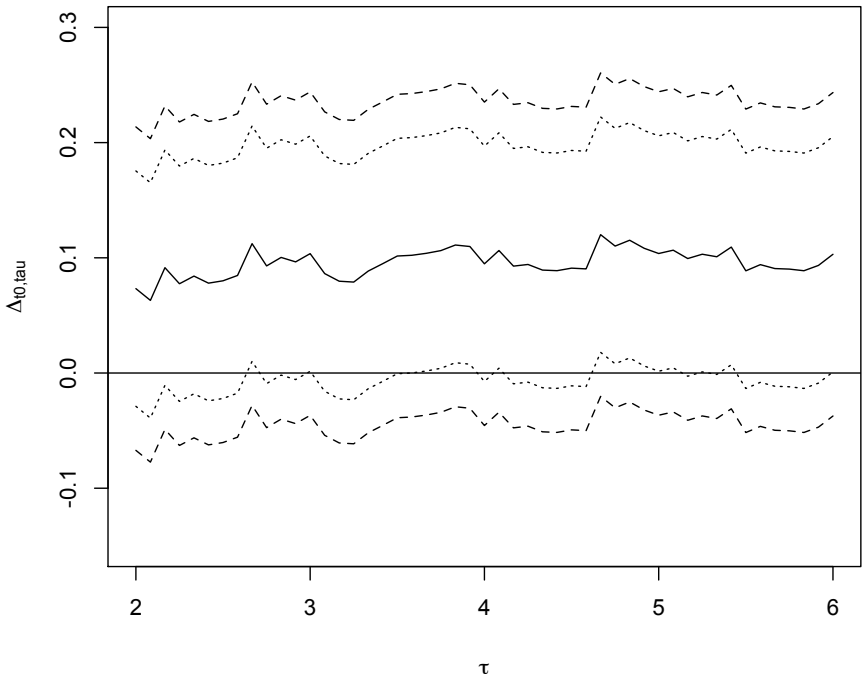

Fig. 2. $\Delta_{t_0,\tau}$ *based on* $\widehat{AUC}^{(CV)(M_2)}_{t_0,\tau} - \widehat{AUC}^{(CV)(M_1)}_{t_0,\tau}$ *(Solid line) for fixed* t_0 *and various values of* τ*where* M_2 *is the model including gene score and* M_1 *is the model not including gene score; pointwise confidence interval (dotted lines) and simultaneous confidence bands (dashed lines).*

References

Begg, C. B. (1991). Advances in statistical methodology for diagnostic medicine in the 1980's. *Statistics in Medicine 10*, 1887–1895.

Bilias, Y., M. Gu, and Z. Ying (1997). Towards a general asymptotic theory for Cox model with staggered entry. *The Annals of Statistics 25*(2), 662–682.

Breslow, N. and D. Clayton (1993). Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association 88*(421), 9–25.

Cai, T. and S. Cheng (2008). Robust combination of multiple diagnostic tests for classifying censored event times. *Biostatistics 9*(2), 216.

Cai, T., S. Cheng, and L. Wei (2002). Semiparametric Mixed-Effects Models for Clustered Failure Time Data. *Journal of the American Statistical Association 97*(458), 514–523.

Cai, T., M. S. Pepe, Y. Zheng, T. Lumley, and N. S. Jenney (2006). The sensitivity and specificity of markers for event times. *Biostatistics 7*(2), 182–97.

Cai, T., L. Tian, H. Uno, S. Solomon, and L. Wei (2010). Calibrating Parametric Subject-specific Risk Estimation. *Biometrika, in press*.

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

Carter, S., A. Eklund, I. Kohane, L. Harris, and Z. Szallasi (2006). A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nature genetics 38*(9), 1043–1048.

Fleming, T. R. and D. P. Harrington (1991). *Counting Processes and Survival Analysis*. Wiley.

Ha, I., Y. Lee, and J. Song (2001). Hierarchical likelihood approach for frailty models. *Biometrika 88*(1), 233–243.

Hanley, H. A. (1989). Receiver Operating Characteristic (ROC) methodology : the state of the art. *Clinical Reviews in Diagnostic Imaging 29*, 307–35.

Heagerty, P. J., T. Lumley, and M. S. Pepe (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics 56*(2), 337–344.

Hjort, N. (1992). On inference in parametric survival data models. *International Statistical Review/Revue Internationale de Statistique 60*(3), 355–387.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime data analysis 1*(3), 255–273.

Hunter, D. (2005). Gene-environment interactions in human diseases. *Nature Reviews Genetics 6*(4), 287–298.

Jewell, N. and J. Nielsen (1993). A framework for consistent prediction rules based on markers. *Biometrika 80*(1), 153.

Kalbfleisch, J. D. and R. L. Prentice (2002). *The Statistical Analysis of Failure Time Data*. John Wiley & Sons.

Klaassen, C. and J. Wellner (1997). Efficient estimation in the bivariate normal copula model: normal margins are least favourable. *Bernoulli 3*(1), 55–77.

McIntosh, M. W. and M. S. Pepe (2002). Combining several screening tests: Optimality of the risk score. *Biometrics 58*(3), 657–664.

Newey, W., D. McFadden, R. Engle, and D. McFadden (1994). Handbook of econometrics.

Oakes, D. and J. Ritz (2000). Regression in a bivariate copula model. *Biometrika 87*(2), 345.

Park, Y. and L. J. Wei (2003). Estimating subject-specific survival functions under the accelerated failure time model. *Biometrika 90*, 717–23.

Pencina, M., R. D'Agostino, R. D'Agostino, and R. Vasan (2008). Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Statistics in Medicine 27*(2), 157–172.

Pepe, M., Z. Feng, Y. Huang, G. Longton, R. Prentice, I. Thompson, and Y. Zheng (2008). Integrating the predictiveness of a marker with its performance as a classifier. *American Journal of Epidemiology 167*(3), 362.

Pepe, M. S., T. Cai, and G. Longton (2005). Combining predictors for classification using the area under the receiver operating characteristic curve. *Biometrics*. , in press.

Pepe, M. S., H. Janes, G. Longton, W. Leisenring, and P. Newcomb (2004). Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *American Journal of Epidemiology 159*(9), 882–90.

Pepe, M. S. and M. L. Thompson (2000). Combining diagnostic test results to increase accuracy. *Biostatistics 1*, 123–40.

Pitt, M., D. Chan, and R. Kohn (2006). Efficient Bayesian inference for Gaussian copula regression models. *Biometrika 93*(3), 537.

Pollard, D. (1990). *Empirical processes: theory and applications*. Institute of Mathematical Statistics.

Shih, J. and T. Louis (1995). Inferences on the association parameter in copula models for bivariate survival data. *Biometrics 51*(4), 1384–1399.

Su, J. Q. and J. S. Liu (1993). Linear combinations of multiple diagnostic markers. *J. Am. Statist. Assoc. 88*, 1350–1355.

Swets, J. A. and R. M. Pickett (1982). *Evaluation of Diagnostic Systems: Methods from Signal Detection theory*. New York: Academy press.

Tian, L., T. Cai, E. Goetghebeur, and L. Wei (2007). Model evaluation based on the sampling distribution of estimated absolute prediction error. *Biometrika 94*(2), 297.

Uno, H., T. Cai, L. Tian, and L. J. Wei (2007). Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association 102*, 527–37.

van de Vijver, M., Y. He, L. van't Veer, and et. al. (2002). A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med. 345*(25), 1999–2009.

van der Sluis, C., J. Kingma, W. Eisma, and H. ten Duis (1997). Pediatric Polytrauma: Short-term and Long-term Outcomes. *The Journal of Trauma: Injury, Infection, and Critical Care 43*(3), 501.

Van Houwelingen, H. (2007). Dynamic prediction by landmarking in event history analysis. *Scandinavian journal of statistics 34*(1), 70.

van't Veer, L., H. Dai, M. van de Vijver, and et. al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature 415*(6871), 530–6.

Weisner, C., G. Thomas Ray, J. Mertens, D. Satre, and C. Moore (2003). Short-term alcohol and drug treatment outcomes predict long-term outcome. *Drug and Alcohol Dependence 71*(3), 281–294.

Wilson, P., R. D'Agostino, D. Levy, A. Belanger, H. Silbershatz, and W. Kannel (1998). Prediction of coronary heart disease using risk factor categories. *Circulation 97*, 1837–47.

Zeng, D. and D. Lin (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal - Royal Statistical Society, Series B Statistical Methodology 69*(4), 507.

Zheng, Y., T. Cai, and Z. Feng (2006). Application of the time-dependent roc curves for prognostic accuracy with multiple biomarkers. *Biometrics 62*, 279–287.