Mu Yue and Jialiang Li*

Improvement Screening for Ultra-High Dimensional Data with Censored Survival Outcomes and Varying Coefficients

DOI 10.1515/ijb-2017-0024

Abstract: Motivated by risk prediction studies with ultra-high dimensional bio markers, we propose a novel improvement screening methodology. Accurate risk prediction can be quite useful for patient treatment selection, prevention strategy or disease management in evidence-based medicine. The question of how to choose new markers in addition to the conventional ones is especially important. In the past decade, a number of new measures for quantifying the added value from the new markers were proposed, among which the integrated discrimination improvement (IDI) and net reclassification improvement (NRI) stand out. Meanwhile, C-statistics are routinely used to quantify the capacity of the estimated risk score in discriminating among subjects with different event times. In this paper, we will examine these improvement statistics as well as the norm-based approach for evaluating the incremental values of new markers and compare these four measures by analyzing ultra-high dimensional censored survival data. In particular, we consider Cox proportional hazards models with varying coefficients. All measures perform very well in simulations and we illustrate our methods in an application to a lung cancer study.

Keywords: diagnostic accuracy improvement, integrated discrimination improvement, net reclassification improvement, C-statistics, varying-coefficient model, ultra-high dimensional data screening

1 Introduction

Survival analysis involves a collection of commonly used statistical methods for the analysis of failure time data such as biological death, mechanical failure, or credit default. In this context, death or failure is also referred to as an “event”. The time-to-event data is usually censored due to the termination of study. One practical objective of survival analysis is to identify the risk factors and quantify their effects on the failure outcome through hazard regression. Accurate risk prediction can be quite useful for patient treatment selection, prevention strategy or disease management in evidence-based medicine. Specifically, we may study how the conditional hazard function of survival time T depends on the observed p -dimensional covariate $\mathbf{X} = \mathbf{x}$

$$h(t | \mathbf{x}) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T < t + \Delta t | T \geq t, \mathbf{X} = \mathbf{x})}{\Delta t}. \quad (1)$$

According to the probabilistic definition, the conditional hazard function may be interpreted as the instantaneous failure rate at time t given a particular covariate profile $\mathbf{x}$. The most popular regression model is Cox proportional hazards (PH) model introduced by Cox [1, 2]

$$h(t | \mathbf{x}) = h_0(t) \exp(\psi(\mathbf{x})), \quad (2)$$

*Corresponding author: Jialiang Li, Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore; Duke-NUS Graduate Medical School, Singapore Eye Research Institute, 1 Third Hospital Ave, Singapore 168751, Singapore, E-mail: STALJ@NUS.EDU.SG

Yue Mu, National University of Singapore, Singapore, Singapore

in which $h_0(t)$ is the baseline hazard. A linear model formulation $\psi(\mathbf{x}) = \mathbf{x}^T \boldsymbol{\beta}$ is usually adopted. See Klein and Moeschberger [3] and references therein for more detailed literature on Cox PH model.

In recent biomedical studies, especially in cancer research, profiling approaches have been extensively conducted, measuring genome-wide mRNA gene expression levels, DNA modifications (e.g., copy number variation – CNV, SNPs), epigenetic regulation (e.g., DNA methylation and histone modifications), post-transcriptional regulations (e.g., microRNA expression) and others. These studies naturally produce massive data and require special statistical methodology. A representative example is The Cancer Genome Atlas (TCGA) in the USA, which provides comprehensive genomic characterization for a cohort of cancer and normal samples. High-dimensional data offer a unique opportunity to more comprehensively describe the etiology and prognosis of important diseases. Correspondingly in statistical literature we have seen a surge of interest in ultra-high dimensionality survival data set where the covariate dimensionality p grows exponentially or non-polynomially fast with sample size n , i.e., $\log(p) = O(n^\alpha)$ for some $\alpha \in (0, 1/2)$. For ultra-high dimensional data, the first step of data analysis is to rank the importance of covariates based on their marginal associations with the outcome variable and screen out unimportant covariates from the ordered list. Only after this necessary screening stage one can proceed to construct a parsimonious regression model applying more sophisticated model building techniques.

For variable screening procedure, Fan and Lv [4] proposed sure independence screening (SIS) based on marginal correlation ranking and also extended SIS to iterative version (ISIS) to handle difficult cases such as when some important predictors are marginally uncorrelated with the response. In order to deal with more complex real data, Fan et al. [5] extended SIS and ISIS to generalized linear models and robust regression; Fan and Song [6] and Fan et al. [7] extended SIS to nonparametric additive models. Moreover, variable screening methods have been quickly adapted in survival analysis with focus on Cox PH model [8–10], and [11] or other regression models [12, 13], and [14]. Recently, He et al. [15], Song et al. [16], and Li et al. [17] proposed non-model based screening methods.

Most statistical methods developed for failure time data assume that covariate effects on the logarithm of the hazard function are linear and the regression coefficients are constant parameters Fleming and Harrington [18] and Andersen et al. [19]. However, true covariate effects can be more complex than the log-linear effect. An important extension of the standard regression model with constant coefficients is the varying-coefficient model. In recent decades, the varying functional effect in survival analysis has been carefully studied by Murphy [20], Cai and Sun [21], Tian et al. [22], Cai et al. [23], and Cai et al. [24], among others. We will consider varying-coefficient Cox model in this paper and aim at providing a novel screening procedure to improve the model performance.

To assess the gain in model development, researchers proposed many improvement statistics in the literature. For example, Pencina et al. [25] proposed the integrated discrimination improvement (IDI) and the net reclassification improvement (NRI), which have drawn much attention in medical research. The IDI is the improvement of the integral of sensitivity and specificity over all possible cut-off values. Extensions of the IDI to survival data have been proposed by Uno et al. [26], Kerr et al. [27] and Chi and Zhou [28]. NRI attempts to quantify how well a new model correctly reclassifies subjects based on comparison between a baseline model and a new model with the added markers. See Pencina et al. [29], Uno et al. [30] and Shi et al. [31] for more discussion. We use such novel improvement statistics in this paper to assess the gain in screening models as well as some standard improvement statistics. We will then use the improvement measures as an effective way to screen ultra-high dimensional biomarkers.

The main contribution of this paper can be summarized as follows. A formal improvement screening methodology is proposed for survival analysis. In particular, the hazards effects of new biomarkers are modelled as varying coefficient functions of possible confounders in the Cox PH model. To the best of our knowledge, for the present setup, improvement screening has not been thoroughly discussed for varying-coefficient Cox PH model yet. Moreover, motivated by the growing importance of biomarker screening problem, in this paper we consider the ultra-high dimensional data setting, which presents a tremendous novel challenge, and, as far as we have reviewed, there is no previous work on this topic. The rest of the paper is organized as follows. In Section 2, we formulate the varying-coefficient Cox PH model and propose

a detailed procedure for our improvement screening methodology. In Section 3, we evaluate the proposed screening procedures through simulation studies and illustrate our methods via application to the lung cancer data set in Section 4. Some discussion is provided in Section 5.

2 Methodology

Many survival analysis data set involves massive number of covariates, such as genetic and genomics studies. Usually low-dimensional confounders are also collected, such as demographical variables, known risk factors, environmental exposures or other variables [32]. Including relevant confounding variables in the model may produce better prediction results. Furthermore, comparison of the results across different studies is easier since most researches on the same response would adjust a similar set of confounders. In our paper, we consider the baseline model with only a confounding variable and add biomarkers to improve the model performance. We present several novel metrics to quantify the improvement from the new markers for prediction of the patients risk and target to screen ultra-high dimensional markers in varying-coefficient models.

2.1 Model and estimation

Let T , C be the random survival time and the random censoring time respectively. In practice we observe $Y = \min\{T, C\}$ and the censoring indicator $\delta = I(T \leq C)$. Denote $\mathbf{X} = (X_1, \dots, X_p)^T$ to be the corresponding p -dimensional covariate vector and U to be the univariate confounder. Our observed data set $\{(\mathbf{x}_i, \delta_i, y_i, u_i) : \mathbf{x}_i \in \mathbb{R}^p, \delta_i \in \{0, 1\}, y_i \in \mathbb{R}^+, u_i \in \mathbb{R}, i = 1, 2, \dots, n\}$ is an independently and identically distributed random sample from $(\mathbf{X}, \delta, Y, U)$. We consider the varying-coefficient Cox PH model:

$$h(t | x_k, u) = h_0(t) \exp(g_{k0}(u) + g_k(u)x_k), \quad k = 1, \dots, p, \quad (3)$$

where x_k is the realization of the k th covariate, $g_{k0}(\cdot)$ and $g_k(\cdot)$ are the unknown intercept and slope functions of the confounder. B-spline smoothing is usually used to estimate the unknown functions and perform very well in practice.

Spline smoothing methods are well known for their success as interpolating polynomials, and their usefulness in providing a smooth basis approximation to a covariate function of unspecified form. We consider the B-spline basis in this paper. Smith [33] employed B-splines in a linear regression problem; Stone and Koo [34] provided a general discussion of the use of B-splines in a covariate function of likelihood-based regression model; Cheng et al. [35] applied B-splines method in analyzing ultra-high dimensional longitudinal data and Xia et al. [12] adopted B-splines in nonparametric accelerated failure time model.

To estimate the unknown functions, we may construct the following partial likelihood for model (3):

$$L = \prod_{i=1}^n \left[\frac{\exp\{g_{k0}(u_i) + g_k(u_i)x_{ik}\}}{\sum_{j \in R(y_i)} \exp\{g_{k0}(u_j) + g_k(u_j)x_{jk}\}} \right]^{\delta_i}, \quad (4)$$

where x_{ik} denote the realization of k th marker for i th individual and $R(y_i)$ is the risk set at time y_i , i.e., the set of individuals alive an instant before time y_i . Let $\mathbf{B}(\mathbf{u}) = (B_1(u), \dots, B_L(u))^T$ be an equispaced B-spline basis, where L is the dimension of the basis. Write $B(u_i) = (B_1(u_i), \dots, B_L(u_i))^T$, $i = 1, \dots, n$. Under appropriate smoothness assumptions, the Curry-Schonberg theorem [36] implies that $g_{k0}(\cdot)$ and $g_k(\cdot)$ can be approximated by some linear combination of $\mathbf{B}(\mathbf{u})$ which optimizes the following objective function:

$$L = \prod_{i=1}^n \left[\frac{\exp\{B^T(u_i)\gamma_{k0} + B^T(u_i)\gamma_k x_{ik}\}}{\sum_{j \in R(y_i)} \exp\{B^T(u_j)\gamma_{k0} + B^T(u_j)\gamma_k x_{jk}\}} \right]^{\delta_i}, \quad (5)$$

where γ_{k0} and γ_k is a vector of length L . Writing $\widehat{\gamma}_{k0}$ and $\widehat{\gamma}_k$ for the maximizer of eq. (5), we attain the B-spline estimator of g_{k0} and g_k as

$$\widehat{g}_{k0}(u) = \mathbf{B}^T(\mathbf{u})\widehat{\gamma}_{k0}, \quad (6)$$

$$\widehat{g}_k(u) = \mathbf{B}^T(\mathbf{u})\widehat{\gamma}_k. \quad (7)$$

An R package **Splines** is available and can be used to implement the above procedure.

2.2 Improvement screening

Our main purpose is to quantify the added value of new markers through marginal screening. Specifically, suppose we have two nested models:

$$M_0 : h(t | u) = h_0(t) \exp(g_0(u)), \quad (8)$$

$$M_k^* : h(t | x_k, u) = h_0(t) \exp(g_{k0}(u) + g_k(u)x_k), \quad k = 1, \dots, p, \quad (9)$$

where M_0 is the baseline (old) model only with a predictor variable U and M_k^* is the new model with both the confounder and the new marker X_k . When $g_k(u) = 0$, M_k^* reduces to M_0 . We present the general improvement screening algorithm as follows:

Improvement Screening Algorithm

Step1: Fit the baseline model M_0 .

Step2: Fit the k th new model M_k^* , $k = 1, \dots, p$.

Step3: Calculate the improvement screening statistics for each of the p markers.

Step4: Sort the p improvement screening statistics in descending order and select the markers corresponding to the top $\lfloor \frac{n}{\log n} \rfloor$ values.

We next provide detailed discussion for four improvement screening procedures.

2.2.1 l^2 -norm

One may consider an approach based on the l^2 -norm which directly measures the overall functional effects of each marker. Define the population l^2 -norm of a function $g(u)$ as:

$$\|g\|^2 \equiv g \cdot g \equiv \langle g, g \rangle \equiv \int |g(u)|^2 du, \quad (10)$$

and empirical l^2 -norm of an estimated function $\widehat{g}_k(u) = (\widehat{g}_k(u_1), \dots, \widehat{g}_k(u_n))$ as:

$$\|\widehat{g}_k(u)\|^2 = \frac{1}{n} (\widehat{g}_k(u_1), \dots, \widehat{g}_k(u_n)) (\widehat{g}_k(u_1), \dots, \widehat{g}_k(u_n))^T. \quad (11)$$

A greater value of this measure may suggest a stronger association between the marker and failure outcome in the presence of confounder U . After sorting the l^2 -norms of all $\widehat{g}_k(u)$, $k = 1, \dots, p$, in a descending order, the corresponding top markers in the ranked list thus reflect strongest hazard contribution towards the censored failure time and improve the baseline model the most.

2.2.2 Harrell's C-statistic

We next consider Harrell's C-statistic proposed by Harrell et al. [37, 38]. It is essentially a rank-correlation measure, motivated by Kendalls tau for censored survival data [39]. In order to assess the added discrimination offered by the addition of a marker to a prediction model, we may calculate the differences of Harrell's C-statistics between two nested models M_k^* and M_0 :

$$\Delta C_k = C_{M_k^*} - C_{M_0}, \quad k = 1, \dots, p. \quad (12)$$

By ranking the statistics ΔC_k in a descending order we expect that markers with greater value should have greater association with survival outcomes since they are more likely to produce concordant risk prediction. Several R packages such as **Hmisc** or **Survival** can be adopted to implement this procedure. We note that C-statistic is also related to the time-dependent diagnostic accuracy of biomarkers. Therefore the improvement based on Harrell's C may be regarded as an improvement over the prediction accuracy as well.

2.2.3 IDI and NRI

The IDI and NRI provide supplementary information over the difference in the areas under the ROC curves. The IDI assesses the improvement in average sensitivity without sacrificing average specificity. Uno et al. [30] define IDI index at time t as

$$IDI_t = E\{\hat{D}(\cdot, t) | T \leq t\} - E\{\hat{D}(\cdot, t) | T > t\}, \quad (13)$$

where the expectation is with respect to $(\cdot, T)$ and $\hat{D}(\cdot, t) = \hat{Pr}(T \leq t | M_k^*) - \hat{Pr}(T \leq t | M_0)$ is the difference of the risk prediction. This boils down to compute the difference of the average probability for events ($T \leq t$) and non-events ($T > t$), and subtract the obtained values for both models to determine how much difference in average probabilities has increased in the new model compared with the old model. If new markers contribute to risk prediction, the first term will be large in the positive direction and the second term will be large in the negative direction; subtracting them produces a large IDI.

To calculate the IDI for each marker, we fit two Cox PH models to a data set, with and without the new marker. Each model yields estimated risk of disease $\hat{D}$ for every individual, event and non-event, in the data set. The estimated risks from the two fitted models are averaged appropriately, and $\widehat{IDI}$ is computed for the data set using eq. (13).

Similarly we may use NRI which is also based on the idea that a valuable new bio-marker will tend to increase predicted risks or risk categories for subjects who develop events and decrease predicted risks or risk categories for those who do not [27]. NRI is another popular index to quantify how well a new model correctly reclassifies subjects as compared to an old model. For survival time outcome, NRI at time t is formulated as [29]:

$$NRI_t = 2\{Pr(\hat{D}(\cdot, t) > 0 | T \leq t) - Pr(\hat{D}(\cdot, t) > 0 | T > t)\}. \quad (14)$$

Further insightful comments about this metric can be found in Pepe et al. [40].

We rank the p covariates based their respective IDI and NRI and useful markers can be obtained from the top of the ranked list. Take note that a time t must be selected by users to calculate the above IDI and NRI metrics, indicating the accuracy increment for assessing the risk prediction at a fixed time point. They should be interpreted differently from the global measures such as the l^2 -norm and the Harrell's C-Statistic. To calculate IDI and NRI, an R package **survIDINRI** is available.

3 Simulation

To examine the performance of the above improvement screening methods, we conduct simulation studies under practical settings. In all simulations, we consider the following true model:

$$h(t_i) = h_0(t_i) \exp \left(\sum_{k=1}^{p^*} g_k(U_i) X_{ik} \right), \quad (15)$$

where $\mathbf{X}_i = (X_{i1}, \dots, X_{ip})$, $i = 1, \dots, n$ are i.i.d. from multivariate normal distribution with mean $\mathbf{0}_p$, and $\text{Cov}(X_{ij}, X_{ik}) = \rho^{|j-k|}$. Both censoring time C and confounder U are generated from uniform distribution on $(0, 10)$. The correlation ρ is set to be 0.8. For the IDI and NRI approaches, we choose the first and third quartiles (corresponding to $\alpha = 0.25$ and $\alpha = 0.50$, respectively) of the observed time. We consider the ultra-high dimensional case with $p = 1000$ and total sample size $n = 400$ and 800 . In our implementation, we adopt cubic spline with one inner knot without intercept. In the first two cases, we consider $p^* = 20$ important variables among 1000 variables; in the latter four cases we increase the number of important variables to 30 and 50. Our purpose is to screen out the important markers from a large set of candidates.

Case I Let $g_0(u) = 1$, $g_k(u) = (3u - 1)^2$ for $k = 1, 5, 9, 13, 17$, $g_k(u) = 3u^3 + 0.5u$ for $k = 2, 6, 10, 14, 18$, $g_k(u) = 3 \sin(2\pi u)$ for $k = 3, 7, 11, 15, 19$ and $g_k(u) = 3 \exp(-4u)$ for $k = 4, 8, 12, 16, 20$.

Case II Let $g_0(u) = 1$, $g_k(u) = (3u - 1)^2$ for $k = 1, 5, 9, 13, 17$, $g_k(u) = 2 \exp(-(3u - 1)^2) + \exp(-4(u - 3)^2)$ for $k = 2, 6, 10, 14, 18$, $g_k(u) = 0.2 \sin(2\pi u) + 0.2 \cos(2\pi u) + 0.3 \sin(2\pi u)^2 + 0.4 \cos(2\pi u)^3 + 0.5 \sin(2\pi u)^3$ for $k = 3, 7, 11, 15, 19$ and $g_k(u) = 3 \exp(-4u)$ for $k = 4, 8, 12, 16, 20$.

Case III Let $g_0(u) = 1$, $g_k(u) = 1.5u$ for $k = 1, 7, 13, 19, 25$, $g_k(u) = 2u$ for $k = 2, 8, 14, 20, 26$, $g_k(u) = 3 \exp(-4u)$ for $k = 3, 9, 15, 21, 27$, $g_k(u) = 2u^3$ for $k = 4, 10, 16, 22, 28$, $g_k(u) = 2 \exp(-(3u - 1)^2) + \exp(-4(u - 3)^2)$ for $k = 5, 11, 17, 23, 29$, $g_k(u) = (3u - 1)^2$ for $k = 6, 12, 18, 24, 30$.

Case IV Let $g_0(u) = 1$, $g_k(u) = (3u - 1)^2$ for $k = 1, \dots, 5$, $g_k(u) = 2 \exp(-(3u - 1)^2) + \exp(-4(u - 3)^2)$ for $k = 6, \dots, 10$, $g_k(u) = 2u^3$ for $k = 11, \dots, 15$ and $g_k(u) = 3 \exp(-4u)$ for $k = 16, \dots, 20$, $g_k(u) = 1.5u^3 + 0.5u$ for $k = 21, \dots, 25$ and $g_k(u) = 2.5u$ for $k = 26, \dots, 30$.

Case V Let $g_0(u) = 1$, $g_k(u) = 1.5u$ for $k = 1, 6, 11, 16, 21, 26, 31, 36, 41, 46$, $g_k(u) = 2u$ for $k = 2, 7, 12, 17, 22, 27, 32, 37, 42, 47$, $g_k(u) = 3 \exp(-4u)$ for $k = 3, 8, 13, 18, 23, 28, 33, 38, 43, 48$, $g_k(u) = 2u^3$ for $k = 4, 9, 14, 19, 24, 29, 34, 39, 44, 49$, $g_k(u) = 2 \exp(-(3u - 1)^2) + \exp(-4(u - 3)^2)$ for $k = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$.

Case VI Let $g_0(u) = 1$, $g_k(u) = 2u$ for $k = 1, 6, 11, 16, 21, 26, 31, 36, 41, 46$, $g_k(u) = 3 \exp(-4u)$ for $k = 2, 7, 12, 17, 22, 27, 32, 37, 42, 47$, $g_k(u) = 3u^3 + 0.5u$ for $k = 3, 8, 13, 18, 23, 28, 33, 38, 43, 48$, $g_k(u) = 2 \exp(-(3u - 1)^2) + \exp(-4(u - 3)^2)$ for $k = 4, 9, 14, 19, 24, 29, 34, 39, 44, 49$, $g_k(u) = (3u - 1)^2$ for $k = 5, 10, 15, 20, 25, 30, 35, 40, 45, 50$.

The selected functions are shown in Figure 1. We design the settings where all of the non-zero functions used to generate data are roughly equally related to the response. Similar settings were adopted by many earlier authors for numerical studies. They represent the practical study where only sparse sets of variables contribute to the response. We have examined other settings with different p values in extensive simulation studies and obtain quite similar findings as we report here. The simulation results are based on 500 independent runs. Besides the four improvement screening methods, we also report the screening results by using the P -value from fitting marginal Cox regression model to each marker. P -value is usually obtained from a Wald test and is commonly used in practice to evaluate the significance of variables. This marginal screening method is widely applied in high-dimensional research.

The first criterion to evaluate the screening performance is the minimum model size which is the smallest number of covariates that we need to include in a model in order to ensure that all the active variables are selected. In Figure 2, we display the box plots of the minimum model size from the simulations. In most cases, the minimum model sizes concentrate around the number of truly important covariates, indicating satisfactory screening order.

Figure 1: The plots of functions $g_k(u)$ defined in simulation for each of the four cases.

In our implementation, as recommended by Fan and Lv [4], Xia et al. [12] and Cheng et al. [35], we keep the top $\lfloor \frac{n}{\log n} \rfloor$ variables in the ranked list and screen out the rest of the covariates. In Table 1, we report the coverage probability that this selected set of top $\lfloor \frac{n}{\log n} \rfloor$ covariates includes all truly important ones. We observe from Table 1 that all screeners perform quite well with the coverage probability close to one. As the number of important covariates increase in Cases III, IV, V and VI, it is slightly harder to select all of them by only keeping the top $\lfloor \frac{n}{\log n} \rfloor$ variables in the ranked list. In general performances of all methods improve as the sample size increases.

Figure 2: The box plots of the minimum model size covers all important covariates.

Table 1: Coverage probability that the set of top $\lfloor \frac{n}{\log n} \rfloor$ covariates after screening includes all important covariates.

Case	n	p^*	IDI ($\alpha = 0.25$)	IDI ($\alpha = 0.50$)	NRI ($\alpha = 0.25$)	NRI ($\alpha = 0.50$)	C-index	l^2 -norm	P-value
I	400	20	1.00	1.00	0.95	0.96	1.00	1.00	1.00
	800	20	1.00	1.00	1.00	1.00	1.00	1.00	1.00
II	400	20	1.00	1.00	0.97	0.99	1.00	1.00	1.00
	800	20	1.00	1.00	1.00	1.00	1.00	1.00	1.00
III	400	30	0.98	0.99	0.80	0.86	0.99	0.99	1.00
	800	30	1.00	1.00	1.00	1.00	1.00	1.00	1.00
IV	400	30	1.00	1.00	0.90	0.83	1.00	1.00	1.00
	800	30	1.00	1.00	1.00	1.00	1.00	1.00	1.00
V	400	50	0.61	0.71	0.11	0.15	0.81	0.58	0.94
	800	50	1.00	1.00	0.96	0.95	1.00	1.00	1.00
VI	400	50	0.56	0.71	0.14	0.15	0.73	0.50	0.98
	800	50	1.00	1.00	0.94	0.96	1.00	1.00	1.00

p^* : number of truly important covariates; C-index: Harrell's C-statistics.

We note that there was little work on improvement screening available in the literature. Though C-index and l^2 -norm are commonly used in practice, screening based on their improvement has not been fully investigated. It is quite appealing to apply improvement screening based on these familiar statistics. We note that IDI and NRI focus on different aspects of accuracy improvement in diagnostic medicine and their results thus should be interpreted differently.

4 Example: lung cancer

Lung cancer is the leading cancer killer in both men and women in the US. About 1 out of 4 cancer deaths are from lung cancer. Each year, more people die of lung cancer than of colon, breast, and prostate cancers combined [41]. Assessing gene expressions in the lung cancer and detection of the relevant gene expressions are thus crucial for lung cancer prevention and treatment. A substantial number of studies have reported the development of gene expression-based prognostic signatures for lung cancer. The ultimate aim of such studies should be the development of well-validated clinically useful prognostic signatures that improve therapeutic decision making beyond current practice standards [42].

Our example was extracted from a large, training-testing, multi-site blinded validation study [43] based on 442 lung adenocarcinomas. The gene expression data in this example were generated by four different laboratories under a common protocol. The same data set has been analyzed by Li et al. [17] and Xie et al. [44]. In this analysis, we adopt age as the potential confounder, as it is strongly associated with end points of interest in occupational epidemiology [45]. After removing cases with missing information, a total of 439 subjects were included in downstream analysis. The median follow-up time is 46 months (range: 0.03 to 204 months) and the median age at diagnosis is 65 years (range: 33 to 87 years). The overall censoring rate is 46.47%. For each subject, the expressions of 22,283 genes are available. Our target is to identify a set of gene expression signatures that are highly predictive in the lung cancer.

We first fit a baseline Cox PH model including only the age covariate whose effect is modelled as an unknown function and then evaluate new models by adding the genes with varying coefficients as described in Section 2. We then quantify the incremental value for each gene by the four proposed improvement screening approaches. For illustration, we also discretized the event time using the first ($\alpha = 0.25$), second ($\alpha = 0.50$) and the third quartile ($\alpha = 0.75$) points in IDI and NRI methods for the short-term, mid-term and long-term prediction, respectively.

In Table 2, we compare the screening results for the top 20 genes obtained by the proposed improvement screening methods and also the P -value based screening approach. We display the IDs of selected genes and the corresponding screening measures along with their 95% confidence intervals calculated by 500 perturbation-resampling. We highlight genes which are simultaneously selected by at least three screeners. Despite certain degree of overlapping, we note that these methods generally produce distinct list of top genes. Depending on the goal and emphasis of the study we may consider choosing appropriate screening statistics. For example, gene 750 achieves the highest IDI at $\alpha = 0.75$ and the third highest NRI value at $\alpha = 0.75$, yielding 14.3% and 35.2% increment from the baseline model, respectively. If the aim is to improve the prediction for long-term survival, we may consider this gene as an important biomarker. The top genes selected by C-statistics give high concordance for risk prediction. For example, gene 7,603 gives an increase of 15.5% in Harrell's C, suggesting that more than 15.5% cases predicted by the new marker would yield risk prediction in the same direction as the failure time. The top genes selected by the l^2 -norm imply a greater contribution towards the hazard rates and all top genes selected by p -value are highly significant.

We also compare our proposed improvement screening approaches to improvement screening without the varying coefficients. In particular, we consider fitting the following two nested models:

$$M_0 : h(t | u) = h_0(t) \exp(\beta_0 \cdot u), \quad (16)$$

Table 2: Summary of top 20 selected genes from different improvement screening methods by fitting varying-coefficient model for the lung cancer data. Estimates (confidence intervals) are given.

rank	$(\alpha = 0.25)$			$IDI(\alpha = 0.50)$			$IDI(\alpha = 0.75)$			$NRI(\alpha = 0.25)$			$NRI(\alpha = 0.50)$			$NRI(\alpha = 0.75)$			C-index		ρ^2 -norm		p-value
	IDI	ID	IDI	ID	IDI	ID	IDI	ID	IDI	ID	IDI	ID	IDI	ID	IDI	ID	IDI	ID	ID	ρ^2 -norm	ID	ρ^2 -norm	
1	13813	0.12 (0.01, 0.49)	750	0.09 (0.05, 0.15)	750	0.14 (0.10, 0.20)	19317	1.33 (0.06, 1.39)	19900	0.89 (0.47, 1.13)	19517	0.75 (0.45, 0.99)	7603	0.16 (0.11, 0.20)	6633	40.5 (37.4, 43.6)	14645	3.0e-14					
2	18354	0.12 (0.00, 0.40)	12798	0.08 (0.03, 0.15)	12352	0.14 (0.07, 0.20)	12378	1.29 (- 0.25, 1.32)	15415	0.88 (0.47, 1.11)	18506	0.74 (0.46, 0.98)	17032	0.14 (0.09, 0.19)	15686	38.3 (35.3, 41.4)	12352	4.2e-14					
3	13176	0.11 (0.00, 0.40)	10896	0.08 (0.04, 0.14)	14645	0.13 (0.07, 0.20)	11344	1.26 (- 0.20, 1.28)	9789	0.88 (- 0.39, 1.04)	750	0.70 (0.47, 0.97)	5135	0.14 (0.09, 0.19)	17059	36.8 (33.7, 39.9)	5135	9.5e-13					
4	12873	0.07 (0.00, 0.49)	3324	0.08 (0.04, 0.14)	19517	0.13 (0.07, 0.19)	19247	1.26 (0.36, 1.32)	20642	0.87 (0.31, 1.03)	15530	0.69 (- 0.24, 0.90)	19517	0.14 (0.09, 0.19)	7100	33.3 (30.2, 36.3)	13950	1.9e-12					
5	2505	0.06 (0.00, 0.42)	5135	0.08 (0.04, 0.14)	5135	0.12 (0.07, 0.19)	166	1.25 (0.11, 1.30)	16921	0.86 (- 0.45, 1.10)	12352	0.68 (0.40, 0.90)	18471	0.14 (0.09, 0.19)	20801	31.2 (28.1, 34.3)	12832	2.9e-12					
6	5241	0.06 (0.00, 0.29)	21963	0.08 (0.04, 0.14)	19924	0.12 (0.07, 0.19)	8140	1.25 (0.27, 1.33)	16128	0.86 (0.47, 1.07)	20987	0.67 (0.25, 0.83)	12352	0.14 (0.09, 0.19)	16128	30.5 (27.4, 33.6)	19924	6.9e-12					
7	4683	0.06 (0.00, 0.32)	3420	0.08 (0.04, 0.14)	12832	0.12 (0.08, 0.18)	14132	1.25 (- 0.19, 1.30)	16961	0.86 (0.57, 1.08)	16763	0.67 (0.10, 0.84)	14645	0.14 (0.09, 0.19)	16986	29.0 (25.9, 32.1)	9422	1.1e-11					
8	13714	0.06 (0.00, 0.35)	19924	0.08 (0.04, 0.15)	1068	0.11 (0.06, 0.17)	21223	1.25 (- 0.37, 1.32)	15853	0.86 (0.25, 1.03)	7603	0.67(0.31, 0.87)	10896	0.14 (0.09, 0.18)	6797	28.8 (25.7, 31.9)	19954	1.6e-11					
9	1938	0.06 (0.00, 0.45)	14645	0.08 (0.04, 0.14)	15705	0.11 (0.07, 0.16)	17857	1.23 (- 0.02, 1.30)	16839	0.86 (0.42, 1.05)	16192	0.66 (0.28, 0.89)	21712	0.13 (0.09, 0.18)	20201	28.5 (25.5, 31.6)	15705	2.3e-11					
10	16792	0.06 (0.00, 0.38)	14744	0.08 (0.04, 0.14)	16942	0.11 (0.06, 0.17)	13113	1.23 (- 0.56, 1.25)	14311	0.84 (0.34, 1.05)	16757	0.66 (- 0.25, 0.85)	750	0.13 (0.09, 0.18)	5364	27.3 (24.2, 30.4)	4782	3.9e-11					
11	10335	0.05 (0.00, 0.43)	9422	0.07 (0.04, 0.13)	6681	0.11 (0.06, 0.16)	14958	1.23 (0.43, 1.31)	18353	0.84 (0.47, 1.08)	17244	0.66 (0.34, 0.90)	16961	0.13 (0.08, 0.18)	14542	26.3 (23.2, 29.4)	16961	4.0e-11					
12	1199	0.04 (0.00, 0.33)	21712	0.07 (0.04, 0.12)	21963	0.11 (0.07, 0.17)	8429	1.22 (0.25, 1.34)	11380	0.84 (- 0.39, 1.00)	9779	0.66 (0.33, 0.90)	16192	0.13 (0.08, 0.18)	11082	25.9 (22.8, 29.0)	7992	6.0e-11					
13	21769	0.04 (0.00, 0.39)	16961	0.07 (0.04, 0.14)	7603	0.11 (0.05, 0.17)	2647	1.22 (0.41, 1.34)	19629	0.83 (- 0.43, 1.05)	3465	0.65 (0.28, 0.90)	15530	0.13 (0.08, 0.18)	16180	24.8 (21.7, 27.9)	21963	6.2e-11					
14	3324	0.04 (0.00, 0.20)	4782	0.07 (0.03, 0.13)	13298	0.11 (0.05, 0.17)	9956	1.22 (- -0.25, 1.26)	8042	0.83 (0.48, 1.05)	12504	0.65 (0.29, 0.89)	11627	0.13 (0.08, 0.18)	14683	24.5 (21.4, 27.6)	750	6.3e-11					
15	16054	0.04 (0.00, 0.27)	12352	0.07 (0.04, 0.13)	9604	0.11 (0.06, 0.17)	8463	1.21 (- 0.46, 1.28)	17065	0.83 (- 0.47, 1.04)	5073	0.64 (0.13, 0.85)	9422	0.13 (0.08, 0.18)	6114	23.7 (20.6, 26.8)	7865	6.4e-11					
16	13063	0.04 (0.00, 0.27)	18365	0.07 (0.04, 0.13)	11627	0.11 (0.05, 0.17)	73	1.20 (- 0.60, 1.28)	16647	0.82 (0.40, 1.00)	11424	0.64 (- 0.29, 0.84)	5311	0.13 (0.08, 0.18)	11151	23.3 (20.3, 26.4)	17032	6.7e-11					
17	18612	0.03 (0.00, 0.33)	5393	0.07 (0.02, 0.17)	5619	0.11 (0.06, 0.16)	11674	1.20 (-0.11, 1.26)	16248	0.82 (0.24, 1.03)	19942	0.64 (0.17, 0.78)	19837	0.13 (0.08, 0.18)	5671	23.3 (20.2, 26.3)	2219	8.0e-11					
18	9025	0.03 (0.00, 0.26)	19954	0.07 (0.03, 0.13)	17438	0.11 (0.06, 0.17)	7815	1.20 (0.05, 1.26)	9922	0.82 (- 0.29, 1.05)	16996	0.64 (0.27, 0.87)	11759	0.13 (0.08, 0.18)	7398	23.0 (19.9, 26.1)	6737	8.2e-11					
19	4383	0.03 (0.00, 0.36)	12832	0.07 (0.03, 0.14)	18471	0.10 (0.05, 0.17)	20611	1.20 (- -0.01, 1.28)	13344	0.82 (- 0.51, 1.00)	10503	0.64 (- 0.05, 0.83)	13413	0.13 (0.08, 0.18)	17074	21.4 (18.3, 24.5)	14965	8.5e-11					
20	5058	0.03 (0.00, 0.26)	15197	0.07 (0.03, 0.13)	21426	0.10 (0.05, 0.16)	21419	1.19 (0.30, 1.31)	5184	0.81 (- 0.40, 0.99)	19900	0.64 (0.20, 0.84)	21383	0.13 (0.08, 0.17)	20064	21.4 (18.3, 24.5)	9631	8.8e-11					

ID: gene ID; C-index: Harrell's C-statistics.

$$M_k^* : h(t | x_k, u) = h_0(t) \exp(\beta_{k0} \cdot u + \beta_{k1} \cdot x_k + \beta_{k2}(u \cdot x_k)), \quad k = 1, \dots, p, \quad (17)$$

where M_0 is the baseline model only with a confounder variable U in a parametric form and M_k^* is the k th new model with the additional marker X_k on top of the baseline. The same improvement screening statistics are implemented and the results are shown in Table 3. In general IDI, NRI and the improvement in Harrell's C values in Table 3 are smaller than those in Table 2. For example, at $\alpha = 0.25$, the first gene selected under IDI using the varying-coefficient attains an IDI 0.12 while that using the constant coefficient achieves the highest IDI 0.10. Such a comparison may suggest that the varying-coefficient model could lead to relatively stronger improvement for the hazards regression than the constant coefficient model. Using more sophisticated semi-parametric varying-coefficient model may provide more accurate model specification for a high-dimensional analysis.

We next report the Pearson's correlation matrix for these measurements in Table 4. The correlation indicates the degree of pairwise agreement between two quantitative improvement statistics. For example, the correlation between NRI and IDI is computed using the following sample statistic:

$$\frac{\sum_i^n = 1(NRI_i - \bar{NRI})(IDI_i - \bar{IDI})}{\sqrt{\sum_i^n = 1(NRI_i - \bar{NRI})^2 \sum_i^n = 1(IDI_i - \bar{IDI})^2}},$$

where NRI_i and IDI_i are the estimated NRI and IDI for the i th marker and $\bar{NRI}$ and $\bar{IDI}$ are the sample means of NRI and IDI, respectively. We observe that the degree of linear correlation between Harrell's C-Statistics & p -value is relatively high with Pearson's correlation 0.619. This may suggest a moderate agreement between these two methods. However, the degree of linear correlation between l^2 -norm and other measures are relatively low. Furthermore, there is a relative high agreement between IDI & NRI (or Harrell's C-Statistics) and between NRI & Harrell's C-Statistics for $\alpha = 0.5$ and $\alpha = 0.75$ but relatively low agreement for $\alpha = 0.25$. Since these methods may address different types of improvement, we thus expect such an empirical disagreement.

We display the estimated varying-coefficient functions $g_k(\text{age})$ for the top three genes in the ranked list by methods IDI($\alpha = 0.25$), IDI($\alpha = 0.75$), NRI($\alpha = 0.25$) and NRI($\alpha = 0.75$) in Figure 3. All of the functions are quite different from a straight line. Source code for this data analysis is available at <http://www.stat.nus.edu.sg/~stalj/>.

5 Discussion

We intend to provide a practical screening approach when the goal is to seek markers with the greatest improvement over existing risk prediction models. Besides the four approaches we examine in this paper, there are additional improvement screening procedures available such as the improvement in odds ratios, hazards ratios, P-values and other parametric and nonparametric measures. They may also be implemented in a similar way. More theoretical development is needed to justify the theoretical properties such as the sure screening properties and the control of false discovery rates.

We also want to highlight the importance of considering sophisticated nonparametric modelling techniques such as the spline estimates of varying coefficients. Many biological effects are nonlinear and a simple description using parametric methods may be misleading and non-informative. Nonparametric and semi-parametric regression methods are well developed in theory and implementation and could serve as useful alternative choices for practitioners.

The method for varying coefficient function is limited to a scalar confounder U . Many applications may involve a set of confounders such as age, gender and other low-dimensional variables. It is usually difficult to smooth more than one index variables in the varying coefficient due to the curse of dimensionality. To address such a concern one may need to adopt certain functional structure for the multiple confounder effects. Model selection and model validation are necessary development for such a goal.

Table 3: Summary of top 20 selected genes from different improvement screening methods by fitting constant coefficient model for the lung cancer data. Estimates (confidence intervals) are given.

rank	IDI($\alpha = 0.25$)			IDI($\alpha = 0.50$)			IDI($\alpha = 0.75$)			NRI($\alpha = 0.25$)			NRI($\alpha = 0.50$)			NRI($\alpha = 0.75$)			C-index			ℓ^2 -norm	
	ID	IDI		ID	IDI		ID	IDI		ID	NRI		ID	NRI		ID	NRI		ID	C-index	ID	ℓ^2 -norm	
1	13813	0.10 (0.00,0.36)	14645	0.08 (0.04,0.14)	12352	0.14 (0.08,0.20)	8140	1.25 (0.59,1.31)	19900	0.93 (0.49,1.09)	19517	0.75 (0.43,0.97)	7603	0.15 (0.10,0.20)	6797	28.7 (26.8,30.5)							
2	7967	0.02 (0.00,0.08)	12798	0.08 (0.03,0.13)	14645	0.13 (0.07,0.19)	21223	1.25 (0.17,1.31)	9789	0.88 (0.48,1.05)	18506	0.73 (0.47,0.99)	17032	0.14 (0.09,0.19)	16128	25.0 (23.2,26.9)							
3	13617	0.01 (0.00,0.07)	5135	0.07 (0.04,0.14)	750	0.13 (0.08,0.18)	19317	1.24 (0.69,1.35)	6895	0.88 (0.33,1.02)	750	0.72 (0.47,0.95)	14645	0.14 (0.09,0.19)	11082	23.6 (21.8,25.5)							
4	12035	0.01 (0.00,0.10)	12832	0.07 (0.03,0.13)	19517	0.13 (0.07,0.19)	19003	1.22 (0.32,1.31)	16961	0.88 (0.55,1.07)	11759	0.69 (0.31,0.89)	12352	0.14 (0.09,0.19)	6633	22.3 (20.5,24.1)							
5	15556	0.01 (0.00,0.09)	12352	0.07 (0.04,0.12)	12832	0.12 (0.07,0.17)	19247	1.22 (0.52,1.29)	16921	0.87 (0.33,1.09)	15530	0.69 (0.19,0.94)	10896	0.14 (0.09,0.19)	5364	20.5 (18.6,22.3)							
6	15889	0.01 (0.00,0.06)	19924	0.07 (0.03,0.13)	19924	0.11 (0.06,0.17)	5331	1.21 (0.33,1.28)	15853	0.87 (0.39,1.03)	14645	0.68 (0.31,0.91)	19517	0.14 (0.09,0.19)	20064	20.0 (18.1,21.8)							
7	16030	0.01 (0.00,0.06)	21963	0.07 (0.03,0.12)	5135	0.11 (0.06,0.18)	73	1.20 (0.92,1.29)	19629	0.87 (0.45,1.06)	20494	0.66 (0.40,0.87)	18471	0.14 (0.09,0.19)	7398	19.6 (17.7,21.4)							
8	2229	0.01 (0.00,0.04)	9631	0.07 (0.04,0.10)	15705	0.11 (0.06,0.16)	10844	1.20 (1.18,1.23)	15415	0.87 (0.50,1.12)	21972	0.66 (0.25,0.92)	21712	0.14 (0.09,0.18)	11151	18.0 (16.2,19.8)							
9	4782	0.01 (0.00,0.03)	16030	0.07 (0.03,0.12)	5619	0.10 (0.06,0.16)	7815	1.20 (0.16,1.27)	7557	0.86 (0.43,1.04)	10863	0.65 (0.36,0.86)	16961	0.14 (0.09,0.18)	21587	17.4 (15.6,19.2)							
10	8604	0.01 (0.00,0.08)	750	0.07 (0.03,0.12)	16942	0.10 (0.06,0.17)	15058	1.20 (0.69,1.30)	18365	0.86 (0.53,1.05)	12504	0.65 (0.32,0.86)	15530	0.13 (0.09,0.18)	7104	15.6 (13.8,17.5)							
11	13644	0.01 (0.00,0.04)	9422	0.07 (0.04,0.10)	21963	0.10 (0.06,0.15)	18000	1.20 (0.28,1.26)	11380	0.85 (0.52,1.06)	2662	0.65 (0.34,0.86)	9422	0.13 (0.09,0.18)	15020	15.1 (13.3,17.0)							
12	21383	0.01 (0.00,0.03)	4782	0.07 (0.03,0.13)	13298	0.10 (0.05,0.16)	459	1.19 (0.11,1.26)	16128	0.85 (0.45,1.08)	16192	0.65 (0.35,0.90)	11759	0.13 (0.08,0.18)	16082	14.5 (12.7,16.4)							
13	4773	0.01 (0.00,0.03)	13950	0.07 (0.03,0.11)	21383	0.10 (0.04,0.16)	8429	1.19 (0.99,1.30)	11230	0.84 (0.53,1.05)	17438	0.65 (0.36,0.87)	5135	0.13 (0.08,0.18)	10537	13.4 (11.5,15.2)							
14	5133	0.01 (0.00,0.06)	16961	0.07 (0.04,0.10)	18471	0.10 (0.04,0.17)	8476	1.19 (0.08,1.25)	20089	0.84 (0.50,1.06)	12352	0.64 (0.40,0.87)	21144	0.13 (0.08,0.18)	16493	13.2 (11.3,15.0)							
15	5135	0.01 (0.00,0.03)	2219	0.06 (0.03,0.11)	10155	0.10 (0.05,0.16)	2249	1.17 (0.34,1.23)	7313	0.83 (0.50,1.02)	16034	0.64 (0.11,0.86)	7992	0.13 (0.08,0.18)	15417	13.1 (11.3,15.0)							
16	6681	0.01 (0.00,0.03)	15705	0.06 (0.02,0.12)	19954	0.10 (0.05,0.16)	18550	1.17 (0.30,1.26)	10896	0.83 (0.55,1.04)	15091	0.64 (0.29,0.85)	1349	0.13 (0.08,0.18)	14656	12.9 (11.1,14.8)							
17	17077	0.01 (0.00,0.06)	3823	0.06 (0.03,0.13)	7865	0.10 (0.05,0.16)	14958	1.17 (0.79,1.25)	13344	0.83 (0.48,1.00)	16194	0.64 (0.21,0.79)	750	0.13 (0.08,0.18)	10837	12.9 (11.0,14.7)							
18	19658	0.01 (0.00,0.08)	5619	0.06 (0.02,0.12)	16961	0.10 (0.06,0.16)	15530	1.16 (0.29,1.27)	16839	0.83 (0.38,1.08)	10059	0.64 (0.18,0.85)	13413	0.13 (0.08,0.18)	7100	12.8 (11.0,14.7)							
19	19882	0.01 (0.00,0.04)	10896	0.06 (0.03,0.12)	17438	0.10 (0.05,0.16)	21419	1.16 (0.51,1.24)	10484	0.82 (0.53,1.00)	14175	0.64 (0.25,0.86)	16942	0.13 (0.08,0.18)	20174	12.2 (10.4,14.0)							
20	21832	0.01 (0.00,0.06)	19954	0.06 (0.03,0.11)	18506	0.10 (0.06,0.14)	19495	1.15 (0.33,1.25)	20642	0.82 (0.40,1.01)	14183	0.64 (0.36,0.89)	21383	0.13 (0.08,0.18)	20082	12.1 (10.3,13.9)							

IDI: gene ID; C-index: Harrell's C-statistics; ℓ^2 -norm: $\|\hat{\beta}_{k1} + \hat{\beta}_{k2}u\|^2$.

Table 4: Pearson’s correlation matrix for screening measurements.

	IDI($\alpha = 0.25$)	NRI($\alpha = 0.25$)	C-index	I^2 -norm	p-value
IDI($\alpha = 0.25$)	1.000	0.360	0.338	0.179	0.148
NRI($\alpha = 0.25$)	0.360	1.000	0.303	0.059	0.106
C-index	0.338	0.303	1.000	0.259	0.619
I^2 -norm	0.179	0.059	0.259	1.000	0.149
p-value	0.148	0.106	0.619	0.149	1.000
	IDI($\alpha = 0.50$)	NRI($\alpha = 0.50$)	C-index	I^2 -norm	p-value
IDI($\alpha = 0.50$)	1.000	0.702	0.837	0.291	0.489
NRI($\alpha = 0.50$)	0.702	1.000	0.758	0.179	0.453
C-index	0.837	0.758	1.000	0.259	0.619
I^2 -norm	0.291	0.179	0.259	1.000	0.149
p-value	0.489	0.453	0.619	0.149	1.000
	IDI($\alpha = 0.75$)	NRI($\alpha = 0.75$)	C-index	I^2 -norm	p-value
IDI($\alpha = 0.75$)	1.000	0.717	0.837	0.272	0.479
NRI($\alpha = 0.75$)	0.717	1.000	0.659	0.159	0.361
C-index	0.837	0.659	1.000	0.259	0.619
I^2 -norm	0.272	0.159	0.259	1.000	0.149
p-value	0.479	0.361	0.619	0.149	1.000

Figure 3: Estimated function of $g_k(\text{age})$ for the top three gene expressions in the ranked list by IDI($\alpha = 0.25$), IDI($\alpha = 0.50$), IDI($\alpha = 0.75$), NRI($\alpha = 0.25$), NRI($\alpha = 0.50$) and NRI($\alpha = 0.75$) using cubic spline with one inner knot and without intercept.

Acknowledgment: We thank the Editor, the Associate Editor and the referees for helpful comments.

Funding: Li's work was partially supported by National Medical Research Council in Singapore and AcRF R-155-000-174-114.

References

1. Cox DR. Regression models and life tables (with discussion). *J R Stat Soc* 1972;34:187–220.
2. Cox DR. Partial likelihood. *Biometrika* 1975;62:269–76.
3. Klein JP, Moeschberger ML. *Survival analysis: techniques for censored and truncated data*. New York: Springer Science & Business Media, 2005.
4. Fan J, Lv J. Sure independence screening for ultrahigh dimensional feature space. *J R Stat Soc Ser B (Stat Method)* 2008;70:849–911.
5. Fan J, Samworth R, Wu Y. Ultrahigh dimensional feature selection: beyond the linear model. *J Mach Learn Res* 2009;10:2013–38.
6. Fan J, Song R. Sure independence screening in generalized linear models with np-dimensionality. *Ann Stat* 2010; 38:3567–604.
7. Fan J, Feng Y, Song R. Nonparametric independence screening in sparse ultra-high-dimensional additive models. *J Am Stat Assoc* 2011;106(494):544–557.
8. Fan J, Li R. Variable selection for Cox's proportional hazards model and frailty model. *Ann Stat* 2002;30:74–99.
9. Fan J, Feng Y, Wu Y. High-dimensional variable selection for Cox's proportional hazards model. In *Borrowing strength: theory powering applications—a Festschrift for Lawrence D. Brown*. New York, USA: Institute of Mathematical Statistics, 2010:70–86.
10. Bradic J, Fan J, Jiang J. Regularization for Cox's proportional hazards model with np-dimensionality. *Ann Stat* 2011;39:3092.
11. Zhao SD, Li Y. Principled sure independence screening for Cox models with ultra-high-dimensional covariates. *J Multivariate Anal* 2012;105:397–411.
12. Xia X, Jiang B, Li J, Zhang W. Low-dimensional confounder adjustment and high-dimensional penalized estimation for survival analysis. *Lifetime Data Anal* 2016;22(4):547–569.
13. Huang J, Ma S, Xie H. Regularized estimation in the accelerated failure time model with high-dimensional covariates. *Biometrics* 2006;62:813–20.
14. Johnson BA, Lin D, Zeng D. Penalized estimating functions and variable selection in semiparametric regression models. *J Am Stat Assoc* 2008;103:672–80.
15. He X, Wang L, Hong HG. Quantile-adaptive model-free variable screening for high-dimensional heterogeneous data. *Ann Stat* 2013;41:342–69.
16. Song R, Lu W, Ma S, Jeng XJ. Censored rank independence screening for high-dimensional survival data. *Biometrika* 2014;101(4):799–814.
17. Li J, Zheng Q, Peng L, Huang Z. Survival impact index and ultrahigh-dimensional model-free screening with survival outcomes. *Biometrics* 2016;72(4):1145–1154.
18. Fleming TR, Harrington DP. *Counting processes and survival analysis*, vol. 169, New Jersey: John Wiley & Sons, 2011.
19. Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical models based on counting processes*. New York: Springer Science & Business Media, 2012.
20. Murphy SA. Testing for a time dependent coefficient in Cox's regression model. *Scand J Stat* 1993;20(1):35–50.
21. Cai Z, Sun Y. Local linear estimation for time-dependent coefficients in Cox's regression models. *Scand J Stat* 2003;30:93–111.
22. Tian L, Zucker D, Wei L. On the Cox model with time-varying regression coefficients. *J Am Stat Assoc* 2005;100:172–83.
23. Cai Z, Fan J, Li R. Efficient estimation and inferences for varying-coefficient models. *J Am Stat Assoc* 2000;95:888–902.
24. Cai J, Fan J, Zhou H, Zhou Y. Hazard models with varying coefficients for multivariate failure time data. *Ann Stat* 2007;324–54.
25. Pencina MJ, D'Agostino RB, Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008;27:157–72.
26. Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei L. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med* 2011;30:1105–17.
27. Kerr KF, McClelland RL, Brown ER, Lumley T. Evaluating the incremental value of new biomarkers with integrated discrimination improvement. *Am J Epidemiol* 2011;174:364–74.
28. Chi YY, Zhou XH. The need for reorientation toward cost-effective prediction: Comments on 'Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond' by Pencina et al., *Statistics in Medicine* (DOI: 10.1002/sim.2929). *Stat Med* 2008;27:182–4.
29. Pencina MJ, D'Agostino RB, Steyerberg EW. Extensions of net reclassification improvement calculations to measure usefulness of new biomarkers. *Stat Med* 2011;30:11–21.

30. Uno H, Tian L, Cai T, Kohane IS, Wei L. A unified inference procedure for a class of measures to assess improvement in risk prediction systems with survival data. *Stat Med* 2013;32:2430–2442.
31. Shi H, Cheng Y, Li J. Assessing diagnostic accuracy improvement for survival or competing-risk censored outcomes. *Canadian J Stat* 2014;42:109–25.
32. Cheng MY, Zhang W, Chen LH. Statistical estimation in generalized multiparameter likelihood models. *J Am Stat Assoc* 2009;104(487):1179–1191.
33. Smith PL. Hypothesis testing in B-spline regression. *Commun Stat Simul Comput* 1982;11:143–57.
34. Stone CJ, Koo CY. Additive splines in statistics. *Proc Stat Comp Sec Am Stat Assoc* 1985;27:45–8.
35. Cheng MY, Honda T, Li J, Peng H. Nonparametric independence screening and structure identification for ultra-high dimensional longitudinal data. *Ann Stat* 2014;42:1819–49.
36. Curry HB, Schoenberg IJ. On Pólya frequency functions IV: the fundamental spline functions and their limits. *J d'anal Math* 1966;17:71–107.
37. Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982;247:2543–6.
38. Harrell FE, Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984;3:143–152.
39. Brown BW, Hollander M, Korwar RM. Nonparametric tests of independence for censored data with application to heart transplant studies. Technical report, DTIC Document, 1973.
40. Pepe M, Feng Z, Gu J. Comments on 'Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond' by MJ Pencina et al., *Statistics in Medicine* (DOI:10.1002/sim.2929). *Stat Med* 2008;27:173–81.
41. Lu Y, Lemon W, Liu PY, Yi Y, Morrison C, Yang P, et al. A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. *PLoS Med* 2006;3:e467.
42. Subramanian J, Simon R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Nat Cancer Inst* 2010;102:464–74.
43. Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008;14:822–7.
44. Xie Y, Xiao G, Coombes KR, Behrens C, Solis LM, Raso G, et al. Robust gene expression signature from formalin-fixed paraffin-embedded samples predicts prognosis of non-small-cell lung cancer patients. *Clin Cancer Res* 2011;17:5705–14.
45. Consonni D, Bertazzi PA, Zocchetti C. Why and how to control for age in occupational epidemiology. *Occup Environ Med* 1997;54:772–6.